



Novel strategies for control of fermentation processes

Mears, Lisa

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Mears, L. (2016). *Novel strategies for control of fermentation processes*. Technical University of Denmark.

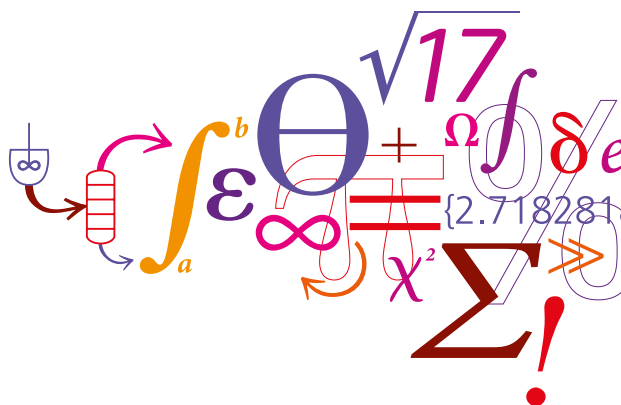
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Novel strategies for control of fermentation processes



Lisa Mears

PhD Thesis

March 2017



Technical University of Denmark

Novel strategies for control of fermentation processes

PhD thesis

Lisa Mears

Department of Chemical and Biochemical Engineering
Technical University of Denmark

Supervisors

Krist V. Gernaey
Gürkan Sin
Stuart M. Stocks
Benny Cassells

Copyright©: Lisa Mears

March 2017

Address: CAPEC-PROCESS Research Center
Department of Chemical and Biochemical Engineering
Technical University of Denmark
Building 229
Dk-2800 Kgs. Lyngby
Denmark

Phone: +45 4525 2800

Web: www.capec-process.kt.dtu.dk/

Print: STEP

DTU Chemical Engineering
Department of Chemical and Biochemical Engineering

Abstract

There is increasing interest in applying more advanced control strategies to biological processes in order to optimise the operation of these complex systems. In the past years, the major increases in product titre have been achieved mainly by genetic engineering approaches, which has lead to highly optimised industrial host strains. The focus of this project is instead on engineering of the process. The question to be answered in this thesis is, given a highly optimised industrial host strain, how can we operate the fermentation process in order to maximise the productivity of the system?

In order to develop control strategies a significant effort must be invested into developing process models and establishing process understanding. Both data-driven modelling and mechanistic modelling approaches are considered in this work. Firstly, multivariate analysis is applied to production scale data from Novozymes A/S in order to predict the product concentration which is measured at the end of the batch. This is achieved with an average prediction error of 7.4%. The purpose of developing the model, is mainly in order to identify key process parameters which show variance relevant to the product concentration, and to identify process trends which lead to higher titres. The application of multivariate methods, in order to provide process insights, creates value from the vast datasets which are collected in industry.

A mechanistic model approach is then considered, based on previous work by Albaek et al (2012). This model describes the fungal processes operated in the fermentation pilot plant at Novozymes A/S. This model is investigated using uncertainty analysis methods in order to assess the applicability to control applications. A mechanistic model approach is desirable, as it is a predictive method which is able to be extrapolated outside of the conditions used to develop the model. For this reason, the mechanistic model approach is further investigated in this work.

The mechanistic model analysis showed that it provided a robust description of the physical system, however there was a relatively high uncertainty in the description of the biological processes. For control applications the model is applied on-line, and therefore it is investigated whether the model prediction may be improved by incorporating available measurement data. A stoichiometric balance approach is applied in order to estimate model parameters including the rate of biomass formation and the rate of product formation. This leads to an increased prediction accuracy in the biological part of the model. The mechanistic model may then be applied as a valuable on-line monitoring tool.

The control strategy development follows on from the on-line model application. The aim of the control strategy is to maximise the total product achieved per batch. There is a demand to maximise the total product in each batch in industry, in order to meet increasing product demands with a limited capacity. The control algorithm is then defined in order to maximise the mass in the system, subject to the oxygen transfer rates in the system. Since the aim is to control to a target fill in a target time, a predictive model-based control algorithm is developed where by the model is simulated to the end of batch time at each model iteration. This provides

a prediction of the future trajectory of the process, so that it is possible to guide the system to the desired target mass. The control strategy is applied on-line at 550L scale in the Novozymes A/S fermentation pilot plant, and the method is challenged with four different sets of process operating conditions. The controller reliably reaches the desired maximum tank fill, with a maximum error of under 5% of the target in eight experimental runs. The product concentration is not affected by the control strategy when compared to batches utilising a reference controller. This method has the benefit of reducing the variance in the final fill, which not only allows for a more reproducible product mass in a batch operation, but also aids downstream process scheduling and resource allocation activities in the industrial setting.

Dansk resume

Der er en stigende interesse i at anvende avancerede kontrol strategier til biologiske processer for at kunne optimere driften af disse komplekse systemer. I de senere år har genteknologi været den primære kilde til højere produkt koncentrationer, hvilket har resulteret i stærkt optimerede værtsstammer. Dette projekt vil i stedet fokusere på opsætningen af processen. Spørgsmålet, der vil blive søgt besvaret i denne afhandling, er: Givet en stærkt optimeret industriel værtsstamme, hvordan kan vi operere fermenteringsprocessen for at maksimere systemets produktivitet?

For at være i stand til at udvikle kontrolstrategier, skal der lægges en betydelig indsats i udviklingen af pålidelige procesmodeller samt at etablere en god procesforståelse. Både data-drevne samt mekanistiske modelleringsmetoder har været anvendt i dette arbejde. Initialt er multivariant analyse anvendt på et datasæt med industriel fermenteringsdata fra Novozymes A/S. Målet var at være i stand til at forudsige produkt koncentrationen, der er målt i slutningen af processen. Dette er opnået med en gennemsnitlig forudsigelsesfejl på 7.4%. Formålet med at udvikle denne model er hovedsagligt at identificere procesparametre, der udviser en høj varians, der er relevant for den endelige produktkoncentration, samt at identificere procesmønstre, der resulterer i høje produktkoncentrationer. Brugen af multivariate metoder til at give procesindsigt åbner dørene for at skabe værdi af de utallige datasæt, der samles i industrien.

En tilgang, baseret på mekanistisk modellering, er siden anvendt, hvilket bygger på tidligere forskning fra Albaek et al (2012). Denne model beskriver den svampe baserede proces, der køres i fermenteringspilothallen på Novozymes A/S. Denne model er undersøgt ved at analysere af model usikkerheden, for at vurdere anvendeligheden i kontrolhenseender. En tilgang baseret på mekanistisk modellering er hensigtsmæssig, da mekanistiske modeller i deres natur er fremskrivende, hvilket åbner for muligheden for at ekstrapolære uden for de procesbetingelser, der blev brugt til udvikling af modellen. Af denne grund, er en tilgang baseret på mekanistisk modellering videre brugt afdækket i denne afhandling.

Analysen af den mekanistiske model viser at den etablerer en robust beskrivelse af det fysiske system, hvorimod usikkerheden vedrørende de biologiske processer er relativt høj. Til proceskontrol anvendes modellen 'on-line', og derfor undersøges det hvorvidt model usikkerheden kan forbedres ved at inkorporere tilgængelig målbar data fra processen. En stokiometrisk balance er anvendt til at estimere en række model parametre, inklusiv hastigheden af biomasse dannelse, samt hastigheden af produkt dannelse. Dette leder til en øget nøjagtighed i procesfremskrivningen af den biologiske del af modellen. Den mekaniske model kan således blive brugt som et værdifuldt 'on-line' monitoringsværktøj.

Udviklingen af en kontrolstrategi er bygget på brugen af 'on-line' monitoringsværktøjet til udvikling af en prædiktiv modelbaseret styrings algoritme. Formålet med kontrol strategien er at maksimere den totale produktivitet opnået per batch. Industrien tilstræber at maksimere produktformationen for hver enkel batch for at imødekomme den stigende produkt efterspørgsel

med begrænset kapacitet. Kontrolalgoritmen er således indstillet på at maksimere massen i systemet underlagt begrænsninger fra hastigheden af iltoverførsel til systemet. Målet er at styre systemet til en defineret fyldningsgrad inden for en afgrænset tid, hvilket løses med en prædiktiv modelbaseret styrings algoritme. Denne model simuleres til slutningen af processen i hver model iteration. Dette resulterer i en fremskrivning af proces udviklingen, og således er det muligt at guide systemet til den pre-definerede fyldningsgrad. Kontrol strategien er anvendt på 'on-line' på 550L fermenteringer i Novozymes pilothal, hvor metoden er testet under fire forskellige procesbetingelser. Med høj pålidelig styrer kontrol algoritmen systemet til den definerede fyldningsgrad med en maksimal afvigelse på under 5%, målt ud fra otte eksperimentale forsøg. Produktkoncentrationen er ikke forringet af den nye kontrolstrategi hvilket ses ud fra sammenligninger med reference eksperimenter. Metoden har den fordel, at den reducerer variansen i den endelige fyldningsgrad, hvilket ikke kun er brugbart i form af en mere reproducerbar proces, men også har en gavnlig effekt på efterfølgende produktionstrin i forhold til ressource allokering samt planlægning.

Preface

This project was conducted as a collaboration between the department of Chemical and Biochemical Engineering at the Technical University of Denmark, and the Fermentation Pilot Plant at Novozymes. There are many people who have contributed to this work, and I have greatly enjoyed working with everyone involved over the past three years.

To Krist V. Gernaey, for his positivity and enthusiasm for the project, which has been fantastic motivation. There has always been a great atmosphere for discussing the work. To Gürkan Sin, who has given a lot of time to discuss the project, in particular the controller development. To Stuart M. Stocks, who I highly value in the project for his open, honest attitude and for so many interesting discussions. I also thank you for all of the time you have taken to discuss and plan the experiments, and provide the industrial insight which is so important in an applied project such as this. To Benny Cassells, who has provided valuable input on this thesis in the final months of the project.

From Novozymes, I would also like to thank Mads O. Albaek, who developed the model which forms the basis of this project. More than this, he has had a great impact on this work, by providing constructive feedback, sharing ideas and opinions, and always being supportive. I would also like to express my gratitude to Karin Nikolajsen who has provided feedback on every publication, every poster and presentation, and taken the time to discuss the project. In addition, I would like to thank all the fermentation operators who have made these results possible. Thank you for sharing your experience with the process, and for your time to monitor and discuss the processes presented in this work.

In addition to my academic supervisors, I would like to express great thanks to Kris Villez, who welcomed us to his group for the external stay. His excellent ability to teach, and provide constructive feedback made for a highly productive start to this project. His attention to detail when preparing the manuscript was also very highly valued.

I would like to thank Daniela Quintanilla for assisting with the lab protocol for sample analysis. Also to many friends at DTU who have been open to discussion of the project, and also provided encouragement in the final months. Thank you for a fun three years.

I would especially like to thank my family for their support, and Rasmus for his encouragement over the past three years.

Contents

Abstract	iii
Dansk resume	v
Preface	vii
Nomenclature	xiii
I Introduction	1
1 Introduction to the project	3
1.1 Structure of the thesis	3
2 Industrial fermentation processes	5
2.1 Industrial fermentation process operation	5
2.2 Industrial fermentation products	6
2.2.1 Enzymes and Novozymes A/S	6
2.3 Industrial fermentation hosts	8
2.4 Bioprocess development	8
3 Literature review of fed-batch fermentation feed rate control strategies	11
3.1 Introduction	11
3.1.1 Definition of the control problem	11
3.1.2 Measured variables	13
3.1.3 Soft sensors	13
3.1.4 Control system design aspects	13
3.2 Control methods review	13
3.2.1 Open-loop control	13
3.2.2 Adaptive control	14
3.2.3 Model predictive control	16
3.2.4 Fuzzy control	17
3.2.5 Artificial Neural Networks	18
3.2.6 Probing control	19
3.2.7 Statistical process control	20
3.3 Conclusions	21
3.3.1 Discussion of control and manipulated variables	21
3.3.2 Discussion of methods	22
3.3.3 Considerations for applying a control strategy in industry	23

II	Modelling	25
4	Data-driven modelling of fermentation processes	27
4.1	Introduction	27
4.1.1	Data-driven modelling of batch processes	28
4.1.2	The challenges of industrial batch process datasets	28
4.2	Methods	29
4.2.1	Time Scaling Methods	29
4.2.2	Functional analysis	30
4.2.3	Data unfolding	32
4.2.4	Mean centring	33
4.2.5	Variable Scaling Methods	33
4.2.6	Multivariate analysis	33
4.2.7	Model validation methods	34
4.3	Development of a methodology for multivariate analysis of batch process data . .	35
4.4	Case study: Production data from Novozymes A/S	37
4.4.1	Results and discussion	37
4.5	Conclusions and future work	42
5	Mechanistic modelling of fermentation processes	45
5.1	Introduction	45
5.1.1	Application of mechanistic models to fermentation processes	46
5.2	Methods	48
5.2.1	Model description	48
5.2.2	Uncertainty analysis	50
5.3	Results and discussion	51
5.3.1	Expert review	51
5.3.2	Bootstrap method	52
5.4	Conclusions and future work	56
III	Monitoring	59
6	Development of a stoichiometric state estimator	61
6.1	Introduction	61
6.1.1	Bioprocess monitoring	62
6.1.2	Soft sensors applied to fermentation systems	62
6.2	Methods	63
6.2.1	Model description	63
6.2.2	Implementation algorithm	67
6.2.3	Off-line sample analysis	67
6.2.4	The historic dataset	68
6.2.5	Validation batches	68
6.2.6	Statistical assessment of model fit	68
6.2.7	Parameter estimation uncertainty	68
6.3	Results and discussion	69
6.3.1	Calibration using historic dataset	69
6.3.2	On-line validation in Novozymes A/S pilot plant	72
6.3.3	Model uncertainty analysis	77
6.4	Conclusions and future work	82

IV	Control	85
7	Control strategy development	87
7.1	Control objectives and motivation	87
7.1.1	Maximising batch fill	88
7.2	Controller development	90
8	Model-based batch planning	93
8.1	Introduction	93
8.2	Methods	93
8.2.1	Mechanistic process model	93
8.2.2	Process operation	94
8.2.3	Process conditions	94
8.2.4	Batch planning methods	94
8.2.5	On-line measurements	95
8.2.6	Off-line sample analysis	95
8.3	Results and discussion	96
8.3.1	Start fill determination	96
8.3.2	Dissolved oxygen controlled experiments	100
8.3.3	Comparison of batch planning results and experimental data	105
8.4	Conclusions and future work	108
9	Model-based control strategy	109
9.1	Introduction	109
9.2	Methods	109
9.2.1	Process operation	109
9.2.2	Process conditions	109
9.2.3	Off-line sample analysis	110
9.2.4	Controller algorithm	110
9.3	Results and discussion	114
9.3.1	Final fill	114
9.3.2	Dissolved oxygen concentration	116
9.3.3	Headspace pressure	119
9.3.4	Product concentration	120
9.4	Conclusions and future work	121
	Overall conclusions and suggestions for future work	123
	Appendices	127
	Supplementary material: Chapter 3	127
	Supplementary material: Chapter 8	134
	Supplementary material: Chapter 9	137
	List of publications	139
	Bibliography	141

Nomenclature

The following nomenclature is valid for this work, unless otherwise specified in the text. For example, in the literature review chapter, other nomenclature may be used, but in this case this is specifically explained in the text.

Roman letters

a	k_{La} equation fitted parameter
b	k_{La} equation fitted parameter
c	k_{La} equation fitted parameter
CER	Carbon dioxide evolution rate (mol/h)
$C1$	Viscosity equation parameter (-)
C	k_{La} equation parameter (-)
D	Impeller diameter (m)
D^2	Second derivative
DO	Dissolved oxygen concentration (mol/kg)
DO_{max}	Maximum DO profile (%)
DO_{min}	Minimum DO profile (%)
$DO_{measured}$	Measured DO (%)
$DO_{profile}$	Dissolved oxygen profile (%)
DO_{set}	Dissolved oxygen setpoint (%)
DO_v	Dissolved oxygen concentration (mol/L)
DO^*	Dissolved oxygen saturation concentration (mol/kg)
DO_v^*	Dissolved oxygen saturation concentration (mol/L)
d	kla equation fitted parameter
F	Feed rate (L/h)
F_0	Nominal value feed controller (L/h)
F_{evap}	Evaporation rate (kg/h)
FO_{limit}	Maximum F/OUR (L/mol)

$F_{profile}$	Feed rate model trajectory (L/h)
$F_{supervisory}$	Feed rate from supervisory layer (L/h)
G	Substrate concentration (g/kg)
HSP	Headspace pressure (barg)
K	Number of basis functions (-)
Kc	Controller proportional gain ((L/h)/%)
k_La	Volumetric oxygen mass transfer coefficient (h^{-1})
k_s	Shear rate constant
M	Mass (kg)
M_{end}	End mass prediction from model trajectory (kg)
M_r	Molecular mass (g/mol)
M_{target}	Target end mass (kg)
$M_{t=0}$	Starting mass (kg)
m_o	Maintenance oxygen consumption ($mol_{oxygen}/g_{biomass}h$)
m_s	Maintenance substrate consumption ($g_{substrate}/g_{biomass}h$)
NH_3	Ammonia addition (mol/h)
N	Stirrer speed (rpm)
n	Number of impellers
OUR	Oxygen uptake rate (mol/h)
O_{2air}	Mole fraction oxygen (mol/mol)
P	Product concentration (g/kg)
P^*	Saturated vapour water pressure (Pa)
$P_{agitation}$	Power dissipated by agitation (kW)
P_{air}	Power dissipated by aeration (kW)
P_{atm}	Atmospheric pressure (bar)
P_{broth}	Total power input to broth (kW)
$P_{electrode}$	Pressure at electrode (bar)
P_g/P_o	Relative power during aeration (-)
P_{head}	Pressure at headspace (bar)
P_o	Impeller power number
P_v	Product concentration (g/L)

PH	Hydrogen content of product (mol H/cmol P)
PN	Nitrogen content of product (mol N/cmol P)
PO	Oxygen content of product (mol O/cmol P)
Po	Impeller power number
Q_{air}	Air flowrate (NL/min)
q_c	Modelled carbon evolution rate (mol/h)
q_g	Modelled substrate uptake rate (mol/h)
q_o	Modelled oxygen uptake rate (mol/h)
q_p	Modelled product formation rate (mol/h)
q_x	Modelled biomass formation rate (mol/h)
q_w	Modelled water formation rate (mol/h)
R	Universal gas constant (J/mol.K)
S	Substrate concentration (g/L)
S_f	Feed substrate concentration (g/L)
S_{feed}	Concentration of feed (g/kg)
T	Temperature (K)
t_{end}	End time (h)
L	Volume (m^3)
v_g	Actual superficial gas velocity (m/s)
v_{gn}	Superficial gas velocity at normal temp. and pressure (m/s)
X	Biomass concentration (g/kg)
X_v	Biomass concentration (g/L)
XH	Hydrogen content of biomass (mol H/cmol X)
XN	Nitrogen content of biomass (mol N/cmol X)
XO	Oxygen content of biomass (mol O/cmol X)
y_{xo}	Yield coefficient (molO ₂ /g X)
y_{xs}	Yield coefficient (g substrate/gX)
Y_{SC}	Observed yield of CO ₂ on substrate (g/g)
Y_{SO}	Observed yield of O ₂ on substrate (g/g)
Y_{SP}	Observed yield of product on substrate (g/g)
Y_{SX}	Observed yield of biomass on substrate (g/g)

Z Liquid height (m)

Greek letters

α_1 Viscosity equation fitted parameter
 β_1 Viscosity equation fitted parameter
 γ_{xo} Stoichiometric coefficient (mol O_2 /g biomass)
 γ_{xs} Stoichiometric coefficient (g substrate/g biomass)
 $\dot{\gamma}$ Shear rate (s^{-1})
 μ Biomass growth rate (h^{-1})
 μ_{app} Apparent viscosity (Pa.s)
 μ_p Plastic viscosity constant (mPa.s)
 ρ Broth density (g/L)
 ρ_F Feed density (g/L)
 τ Shear stress (Pa)
 τ_o Yield stress (Pa)
 Φ Basis function

Abbreviations

ANN Artificial neural network
CER Carbon evolution rate
DO Dissolved oxygen concentration
DOT Dissolved oxygen tension
HSP Headspace pressure
MPC Model predictive control
OTR Oxygen transfer rate
OUR Oxygen uptake rate
PCA Principal component analysis
PCR Principal component regression
PENSSE Penalised sum of squared errors
PLS Projection to latent structures
RMSSE Root mean sum of squared errors
SPC Statistical process control

Part I

Introduction

1 Introduction to the project

This aim of this project is to develop a novel control strategy to maximise the total product achieved by an industrial fermentation process at Novozymes A/S. In order to develop a control strategy, a significant amount of time must be invested into developing process understanding. One way to expand and consolidate process understanding is by developing process models. This thesis therefore covers topics in modelling and monitoring, with a focus on control applications. The result of this thesis is a novel model-based control strategy of industrial relevance which is tested on-line at pilot scale.

1.1 Structure of the thesis

Part I: Introduction

The first part of this thesis introduces the project and the field of study. Industrial fermentation processes are described, and in doing so, the scope of this work is clearly defined, which is focussed on pilot scale, fed-batch filamentous fungus processes for enzyme production. The current state of the art for fermentation control is described in the literature review, which is focussed on feed rate manipulation for fed-batch operation.

Part II: Modelling

Part II of the thesis deals with modelling of fed-batch fermentation processes. Firstly a data-driven method is investigated in order to characterise production scale data from Novozymes A/S. Due to the challenges of dealing with industrial scale batch process data, a novel methodology is proposed in order to aid model development in the future.

A mechanistic approach is also investigated, based on the prior work of Albaek et al. (2011) and Albaek et al. (2012). This pilot scale model is investigated, and additional uncertainty analysis is conducted in order to assess the model limitations. A mechanistic model is desirable as it represents the current system understanding, and in this way consolidates the process knowledge in an industrial environment. It is also more flexible than a data driven approach, and may be applied to multiple systems by adaptation of certain model parameters. Mechanistic models are also more suited to extrapolation outside of the conditions used to develop the model. For these reasons, a mechanistic model approach is preferred for future work in this project.

Part III: Monitoring

The mechanistic model is further developed for on-line application. Since the ultimate goal of this work is to develop a control strategy, which will be applied in real time, it is assessed if the model prediction may be improved by utilising available measurement data. This then leads to a monitoring tool, based on the mechanistic process model, coupled to a parameter estimation which takes inputs from the on-line process. This allows for successful modelling

of key parameters including the biomass concentration and the product concentration. These parameters are otherwise only available off-line, and with time delay on the order of a day.

Part IV: Control

The final section of the thesis describes the development of a novel control strategy, based upon the mechanistic process model. The objective of the control strategy is to maximise the total product achieved in a fed-batch operation, and to achieve this reliably. To this end, the objective of the control strategy is to maximise the mass in the system subject to the oxygen limitations, which are defined by a set of process operating conditions. This represents an industrial problem, which is applicable to large fed-batch processes. The control strategy is successfully implemented on-line in the 550L scale pilot plant facilities at Novozymes A/S.

2 Industrial fermentation processes

A fermentation process may be described as a process where a substrate is converted into a valuable product by a biological system. The biological system is typically a pure bacteria or fungus, which is often genetically modified in order to optimise the expression of the desired product. The typical industrial fermentation process is operated in a mechanically mixed stainless steel fermenter. For aerobic processes, air is supplied at the bottom of the vessel, and the mechanical mixing dissipates the gas, and enhances oxygen transfer from the gas to the liquid phase. A carbon substrate is required, and this may be supplied in bulk at the start of the process, or may be added to the system over time depending on the type of operation employed.

2.1 Industrial fermentation process operation

Batch fermentations are the most simple to operate, where all carbon source and media components are added in bulk at the start of the fermentation, and the batch then runs until carbon source is depleted. The conditions are highly dynamic, with the substrate concentration decreasing over time, and the biomass and product concentrations increasing. This operation has the advantage that it is simple to operate, and the risk of a contamination is greatly reduced by the closed operation. However, the method requires a long downtime for batch turnaround due to the sterilisation requirements. It is also inefficient, with changing substrate concentrations, and does not allow for control of the growth rate or the product formation rates.

An alternative is continuous operation, where feed is added, and the product stream removed, at an equal rate. The aim is to maintain the system at a steady state with high product formation. This can result in a highly productive process, with a comparably low operational cost. It can allow for the production of products which otherwise become catabolised at high concentrations [Villadsen et al., 2011], or may be inhibitory at high concentrations. However, there are operational challenges, especially at industrial scale, as it requires tightly controlled conditions and robust monitoring methods. There may also be scheduling challenges as the downstream operations cannot always be operated continuously. In addition, this long operation demands a genetically stable production host system, and there is also a higher risk of contamination.

A majority of industrial fermentation processes employ a fed-batch operating mode in a stirred tank [Birol et al., 2002, Bodizs et al., 2007]. The first stage of the fermentation is operated in batch mode with a bulk of carbon source to promote biomass accumulation. Once this bulk is depleted, feeding begins in order to supply the system with carbon source for both product formation and biomass growth and maintenance. This allows for significantly greater biomass and product concentrations than batch operation. In addition, processes can be operated for significantly longer meaning that the down-time is reduced in relation to the process time, to increase equipment utilisation. There is a need for improved monitoring and control of the process in order to supply the feed at a suitable rate, and to monitor the tank fill which is con-

tinuously increasing over the process time. This work focusses entirely on fed-batch operation.

2.2 Industrial fermentation products

Industrial fermentation processes are applied for the production of a wide range of industrial products, including alcohols, amino acids, vitamins and enzymes [Deloitte, 2014, Lee and Kim, 2015, Doran, 1995]. Table 2.1 shows a summary of some fermentation products and an example host organism. By far the largest share of the industrial fermentation industry is now attributed to bioethanol, which is 94% of the industry by volume and 87% in terms of value [Deloitte, 2014].

Industrial biotechnology is becoming increasingly attractive due to many desirable process characteristics. Biological processes offer more sustainable production routes for a variety of products as they are low temperature, low pressure systems, which utilise organic and more sustainable raw materials. There is therefore an increasing trend towards chemical production using biotechnological routes, due to sustainability concerns of traditional chemical processes. For example 1,3-propanediol which is traditionally produced using fossil fuels, is now produced commercially by DuPont Tate & Lyle Bio Products by a fermentation process. This consumes 40 percent less energy and reduces greenhouse gas emissions by 20 percent versus petroleum-based propanediol [DuPont Tate & Lyle Bio Products, 2006]. With the current environmental pressures, biological processes become increasingly important as we aim to reduce our dependence on fossil fuels, and reduce our emissions, so it is expected that this list of fermentation products will continue to grow in the coming years.

In addition, fermentation products may be used as replacements to chemical alternatives in order to make other industrial processes more sustainable. As an example, enzymes, produced by fermentation, are applied in a range of industries in order to replace chemical catalysts, or avoid the need for harsh processing conditions. It is an enzyme production process which is considered in this work.

2.2.1 Enzymes and Novozymes A/S

In 1952, Novozymes A/S brought to market the world's first enzyme produced by fermentation. Since then it has grown to be the world's largest enzyme producer, with 48% of the global market for industrial enzymes in 2015 [Novozymes A/S, b]. Novozymes supplies to five major markets, namely household care, food and beverages, bioenergy, agriculture and technical products and pharmaceutical. In these markets, the application of enzymes allows for more sustainable processes, for example through reduced water consumption, reduced energy consumption, and reduced waste generation.

The application of industrial enzyme products, has a beneficial impact to a wide range of industries. For example, enzymes from Novozymes are supplied to the bioenergy market for production of bioethanol, reducing dependency on fossil fuels. In the textile industry, application of enzymes can save 70,000 liters of water and one ton of CO_2 per ton of knitted fabric in comparison to using traditional chemicals [Novozymes A/S, a]. That equals savings of 20-30 liters of water and a reduction of 0.3 kg CO_2 for each t-shirt produced [Novozymes A/S, a]. Novozymes produces a range of household care products, including enzymes for washing powder, which allows improved performance at reduced temperatures [Novozymes A/S, b]. For these reasons, Novozymes A/S is consistently ranked high in sustainability indices, for example the Dow Jones Sustainability index [Novozymes A/S, b].

Table 2.1: A summary of fermentation products, and an example host organism. Produced based on [Doran, 1995, Waites et al., 2001]

Product	Producing strain
Alcohols	
Ethanol	<i>Saccharomyces cerevisiae</i>
Butanol	<i>Clostridium acetobutylicum</i>
Acetone	<i>Clostridium acetobutylicum</i>
Amino acids	
L-Glutamic acid	<i>Corynebacterium glutamicum</i>
L-Lysine	<i>Brevibacterium flavum</i>
L-Tryptophan	<i>Klebsiella aerogenes</i>
Enzymes	
Cellulase	<i>Trichoderma reesei</i>
Protease	<i>Bacillus licheniformis</i>
α -amylase	<i>Bacillus subtilis</i>
Glucose isomerase	<i>Streptomyces olivaceus</i>
Pectinase	<i>Aspergillus niger</i>
Lipase	<i>Candida cylindraceae</i>
Organic acids	
Citric acid	<i>Aspergillus niger</i>
Lactic acid	<i>Lactobacillus delbrueckii</i>
Acetic acid	<i>Acetobacter xylinum</i>
Polymers	
Dextran	<i>Leuconostoc mesenteroides</i>
Xanthan	<i>Xanthomonas campestris</i>
Vitamins	
Vitamin C	<i>Acetobacter suboxydans</i>
Vitamin B2	<i>Eremothecium ashbyii</i>
Vitamin B12	<i>Pseudomonas denitrificans</i>
Antibiotics	
Penicillins	<i>Penicillium chrysogenum</i>
Tetracycline	<i>Streptomyces aureofaciens</i>

2.3 Industrial fermentation hosts

Table 2.1 provides examples of production hosts for industrial fermentation products, which covers both microbial and fungal strains. These two host types have some general characteristics which affect the process operations. Within these groups, each strain will have its own specific characteristics, however some generalities can be made.

Microbial strains grow as single cells which multiply by cellular division, and therefore have faster growth rates and a more simple cellular structure. They are typically spherical or rod shaped, with a size of a few μm in length. Due to both their size, and a strong external cell wall, they are typically highly robust cells which tolerate high shear rates [Doran, 1995]. These benefits are balanced by some limitations to the complexity of the products they can produce, as they lack the ability to carry out some of the post-translational modifications [Villadsen et al., 2011] which are required for many therapeutic products for example.

Fungi may be characterised by their growth form; either single celled yeast or hyphae forming filamentous strains. Yeast strains have a simple spherical structure, and exist as single cells. They reproduce by budding from a parent cell. They are able to produce more complex post-translational modifications to a protein than microbial strains can, for example glycosylation. There are strains available which are generally regarded as safe (GRAS), for example the commonly used *Saccharomyces cerevisiae* [U.S. Food and Drug Administration, 2015].

Filamentous fungal strains are characterised by the morphology of the cells. On a microscopic level the hyphae extensions result in a complex branched morphology. Over the process time, these filaments may become entangled and create a complex structure, resulting in mycelial clumps and aggregates. In some cases the clumps may become dense and form distinct pellets of cells. Due to these complex cellular structures, the rheological behaviour of a filamentous fermentation broth is challenging to characterise [Posch et al., 2013, Quintanilla et al., 2015]. The rheological properties of the broth have implications for the oxygen transfer in the system. This means that many filamentous fungal processes have issues with oxygen transfer, and may be oxygen transfer limited [Posch et al., 2013]. Despite this disadvantage, filamentous fungus strains are desirable due to very high levels of protein secretion [Villadsen et al., 2011]. In this work, the host strain of interest is a filamentous fungus producing extracellular enzymes.

2.4 Bioprocess development

Bioprocess development is an interdisciplinary challenge, which encompasses a range of activities. This includes strain development, process optimisation, and process transfer between scales, as described in Figure 2.2. Many industrial fermentation processes utilise highly engineered production host strains. Major increases in productivity are through genetic engineering of the strain of interest. For example, a review by Cherry and Fidantsef (2003) describes how wild type strains may produce up to 10g/L of enzymes product, however current engineered strains are in excess of 40g/L, and for example cellulases may be produced at concentrations over 100g/L [Cherry and Fidantsef, 2003]. These major improvements are required in order to create commercially viable products. Other important areas of focus for strain development, in addition to product titre, includes strain stability and removal of by-product formation pathways [Lee and Kim, 2015]. A thorough summary of metabolic engineering activities is provided in Lee and Kim (2015).

Once a strain is developed, the task is then to develop a process with suitable growth conditions to maximise the productivity of the strain. This begins with lab scale trials, where

Table 2.2: A summary of operations for development of an industrial fermentation process

Strain development	Lab scale	Pilot scale	Production scale
<ul style="list-style-type: none"> • Host selection • Metabolic pathway construction • By-product removal • Energy and redox requirements • Increased product tolerance • Optimisation of metabolic fluxes • Carbon source • Product extracellular release 	<ul style="list-style-type: none"> • Strain characterisation • Identify process yields • Analyse strain robustness • Media optimisation • Identify optimal operating conditions 	<ul style="list-style-type: none"> • Optimisation of process operating conditions • Feed strategy development • Process monitoring and control • Process modelling • Scale-up studies • Scale-down studies • Collaboration with downstream processing and formulation • Academic studies 	<ul style="list-style-type: none"> • Process optimisation • Quality control • Product testing • Process scheduling • Resource allocation • Process monitoring and control • Equipment testing and maintenance • On-line troubleshooting • Full production process overview

the strain is characterised, to identify the process yields, product titres, and maximum specific growth rates of the organism. At this stage, media is also optimised based on the requirements of the strain, and process optimisation is initiated. A screening study of operating conditions is conducted to determine suitable operating conditions for the organism, in particular temperature and pH of the process. These processes can be conducted in small shake flasks or may be conducted in small scale parallel bioreactors. There is increasing interest in high throughput screening methods (HTPS) in order to increase speed and efficiency of these initial screening activities [Long et al., 2014]. It is also necessary at this stage to assess the robustness of the strain, and the industrial applicability of the strain, for example with respect to foaming, strain morphology or shear sensitivity. It may be possible to identify areas of concern for the future process development, before large scale testing is conducted.

Once a strain has been characterised at lab scale, it will continue to pilot scale testing, where the process knowledge obtained at the lab scale is further developed in order to design a process operation which achieves the desired yields, and has application to production scale. It is desirable that the final production process is optimal, however it is prohibitively expensive to conduct optimisation studies at the full production scale, which may be around 100m³ [Stocks, 2013]. For this reason, pilot scale facilities are important for optimisation and scale-up studies, in order to take a lab scale experiment successfully into a production scale operation. This requires a consideration of the translation of the process parameters between scales. As discussed in Stocks (2013) this scale dependent understanding is equally a matter of scale-down studies, as it is scale-up. If there is an understanding of the limitations of the final production scale process, it is possible to design a pilot scale operation which best represents these operating conditions [Stocks, 2013]. This should aid scale up activities, as the process conditions are comparable.

It is also necessary at pilot scale to develop monitoring and control methods. After the initial strain development, where large increases in titre are obtained, it is the optimisation of process control which allows for further increases in productivity. This includes feed addition algorithms, and control of the process operating conditions. It is this which is the area of focus of this work.

A successful process in the pilot scale is then transferred to a production scale process. Once a production scale process is established there is still continued collaboration between pilot and production scale, in order to ensure the process is running optimally, and to continually develop the process. This is possible for products of industrial biotechnology, which are not subject to the same stringent regulations as pharmaceutical processes, where there are restrictions on the changes which can be made to an established process. There are additional challenges at production scale, especially in multi-product sites, regarding process scheduling, and resource allocation where a production schedule must be planned for both upstream and downstream operations in order to meet the required product quantities. This requires strict time management of all operations, including also downtime and sterilisation, and materials preparation. In order to minimise deviations from the planned schedule, all processes must be tightly monitored to identify issues in a timely manner, and intervene where necessary. There is also a need to identify equipment faults, for example in sensor calibration or measurement errors. This ensures not only that the schedule can be maintained, but that the processes are reproducible, and the product is of equal quality.

There are additional challenges specific to process development activities in an industrial setting. Industrial fermentations are often difficult to replicate across production sites or between facilities as the small operating differences in the equipment affect the way the batches should be optimally run. In addition, batches run in the same facility can also be affected by batch variations in the growth characteristics of a specific cultivation. For these reasons, it is vital to have optimal control of the process to achieve the desired product quality and yields. There is demand therefore to research strategies to continually monitor the performance of a fermentation which is universal across equipment, and to design control systems which maintain the process to an optimal operation. As a majority of the operations are operated as fed-batch, feed rate regulation is of importance for the process optimisation.

3 Literature review of fed-batch fermentation feed rate control strategies

This literature review provides an introduction to the current state of the art for control of fed-batch fermentation processes, and is therefore an introduction to the problem being addressed in this work. This chapter is published in the following article:

A review of control strategies for manipulating the feed rate in fed-batch fermentation processes
Mears L., Stocks S. M., Sin G., Gernaey K. V.
Journal of Biotechnology 245:34-46. (2017)

3.1 Introduction

A majority of industrial fermentation processes are operated in fed-batch mode. In this case, the rate of feed addition to the system is a focus for optimising the process operation, as it directly impacts metabolic activity, as well as directly affecting the volume dynamics in the system. This review covers a range of strategies which have been employed to use the feed rate as a manipulated variable in a control strategy. This provides an introduction to the problem of control for fed-batch processes and defines the current state of the art for this field. The feed rate is chosen as the focus for this review, as it is seen that this variable may be used towards many different objectives depending on the process of interest, the characteristics of the strain, or the product being produced, which leads to different drivers for process optimisation. This review summarises the methods, as well as focusing on the different objectives for the controllers, and the choice of measured variables involved in the strategy. The conclusion includes a summary of considerations for control strategy development. This then forms the basis for this project in fermentation control strategy development.

3.1.1 Definition of the control problem

This review focusses on control methods for fed-batch fermentation processes, specifically for feed rate control. The challenge is not only to maintain the optimal feed rate, but first to identify how the optimal feed rate is defined. This is a complex issue, which is dependent on the process and product. The feed rate applied to a fermentation system is such a central process parameter, as may be seen in Figure 3.1.

The feed rate directly affects the fermentation process in two ways; metabolic changes and volume dynamics. The substrate concentration affects the metabolic rates; namely the growth rate, specific product formation rate, specific by-product formation rates and oxygen uptake rate. In addition, the concentrations of all elements in the system are affected by the changing

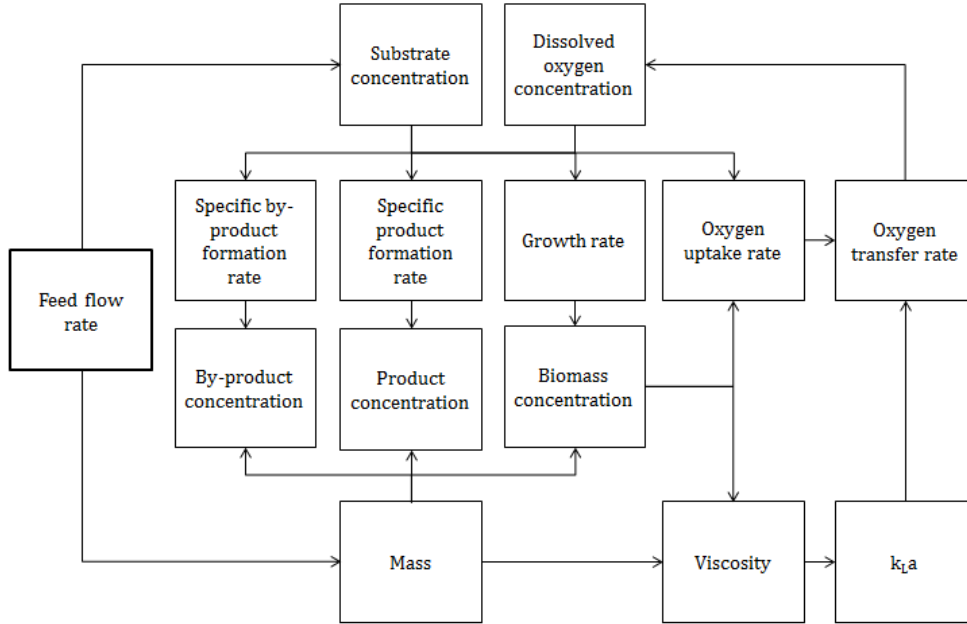


Figure 3.1: Feed flow rate effects in the fed-batch fermentation process.

volume, which is a direct effect of the feed addition. There is also an impact on the oxygen dynamics, through both the oxygen uptake rate, and also the changing oxygen transfer rate, which is partly due to a changing viscosity, as a consequence of a changing biomass concentration. Overall, the feed rate is seen to impact the system in many ways. The control objective is dependent on the process of interest and the economic drivers for the process. Some examples of possible control objectives by manipulating the feed rate include:

- Maximise product concentration
- Minimise by-product formation
- Process yield
- Productivity
- Maximise biomass concentration
- Maintain an oxygen concentration profile

Once a control objective is defined, there are great challenges in how to meet the objective. For example, the final product concentration may be the key driver. In this case, a feed rate must be found which saturates the pathway for product formation. If too little is fed, then the process is not at its maximum productivity. However overfeeding is also a significant problem as this leads to overflow metabolism which can result in by-product formation [Villadsen et al., 2011]. The feed rate which will achieve this optimal point is constantly changing, with non-linear dynamics, due to exponential growth rates, metabolic shifts, volume dynamics, as well as possible feed disturbances. The non-linear dynamics and the disturbances make any control problem implementation challenging in these fed-batch processes.

3.1.2 Measured variables

There are a limited number of probes for industrial scale fermentation applications. The standard probes which are routinely used in industrial scale fermentation allow measurement of the temperature, pH, and dissolved oxygen tension (DO), and in addition there are on-line measurements of the stirrer speed, back pressure and flow rates [Alford, 2006]. Often, off gas composition will also be measured as this provides a vital insight into the metabolic state. Other advanced sensors are available, for example spectroscopic techniques [Cervera et al., 2009], however they are not routinely employed for industrial operation. There is a lack of robust measurements which define key state variables in the fermenter, such as substrate concentration, or biomass concentration.

3.1.3 Soft sensors

A challenge which is often discussed in relation to industrial fermentation technology is the lack of measurements for key variables which define the process [Luttmann et al., 2012, Montague et al., 1989]. A parameter must be robust and without significant time delay if it is to form the basis for a control strategy. Due to the lack of available measurements for key parameters, soft sensors can provide additional control variables for new control methodologies. For example, on-line data can be processed in real-time to calculate the expected value of a variable which is otherwise only measured off-line. This method sees particular applicability in fermentation where a number of key control variables are not measured accurately on-line, or are subject to large time delay [Montague et al., 1989]. Current developments lead to examples of multiple metabolite estimation using soft sensing techniques [Luttmann et al., 2012]. Soft sensing can be a powerful tool for industrial application as part of a framework for QbD and PAT principles since it provides a method for on-line monitoring of the process [Sagmeister et al., 2013, Luttmann et al., 2012].

3.1.4 Control system design aspects

A control strategy is considered to consist of manipulated variables, and controlled variables. In this review, the manipulated variable to be considered will be the feed flow rate. The issue is how to define the controlled variable, since there is a lack of on-line sensors for the variables which would be considered directly applicable to feed rate (such as substrate concentration). As previously discussed, it may be necessary to utilise soft sensors. Control strategies may involve one controlled variable, or in the case of multivariable control, there may be multiple variables controlled using the same controller algorithm. When testing a control strategy it is important to consider system disturbances. For an industrial fermentation process this may include for example batch-to-batch variations in feed concentration, initial biomass concentration and noise in the process measurements.

For each strategy in this review, the manipulated variables, the controlled variables and disturbances considered will be summarised, in addition to the basis for the control strategy. The purpose is to analyse the current state of the art, and discuss the applicability of the control strategies in an industrial context.

3.2 Control methods review

3.2.1 Open-loop control

Many industrial carbon-limited fermentation processes are operated in open loop [Oliveira et al., 2004]. Open loop control may be employed in order to apply a predefined feed rate to the

process, which is based upon the initial conditions for the process and the defined operating conditions. This method requires no on-line measurements, but as a result is not able to reject any disturbances to the system. Predetermined exponential feed profiles are often discussed for the initial growth phase in a process [Henes and Sonnleitner, 2007, Jenzsch et al., 2006a, Wechselberger et al., 2012], where an exponential profile can be calculated based on the initial conditions, and strain specific parameters such as the maximum specific growth rate (μ_{max}). In the work of Jenzsch et al. (2006) this method is used for the purpose of reducing variability between batches caused primarily by variations in initial biomass concentration [Jenzsch et al., 2006a]. Equation 3.1 is applied, where $\mu_{set} \ll \mu_{max}$ such that the batches are reproducible by the end of this exponential growth phase, since the μ_{set} restriction counteracts the disturbance in starting biomass concentration. By operating $\mu_{set} \ll \mu_{max}$ max this achieves the goal of the study which is reproducibility, however it should be noted that it is not generally favourable to operate at this point. In Equation 3.1, Y_{XS} is the yield of biomass on substrate.

$$F(t) = \frac{(\mu_{set} X_0 M_0)}{(Y_{XS}(S_f - S_0))} e^{t\mu_{set}} \quad (3.1)$$

In open loop control there is no adaptation to disturbances. This may be especially problematic for feed rate control strategies, where any batch-to-batch variation in feed concentration is not accounted for. The benefit of the method is however the simplicity of its application and the fact that there is no reliance on measured variables.

3.2.2 Adaptive control

Adaptive control strategies are non-linear controller algorithms, where the controller parameters are adapted over the operation. Adaptive strategies cover a wide range of methods, but the common element is that certain parameters of the controller change in order to better respond to non-linear dynamics or uncertainties in the operation of a system. The method used for this adaption separates the controller types. Due to non-linear dynamics, adaptive strategies are interesting for fermentation processes [Smets et al., 2004].

Gain scheduling is one method of adaptive control. In this case, the controller gain is not constant, but adapted based on prior knowledge of the system, in order to account for changing dynamics. The pre-programmed tuning, can be carried out based on data analysis from past batches and analysis of the response of the controller over time. Hisbullah and Ramachandran (2002) applied gain scheduling to the feed rate control for a simulated *S. cerevisiae*, as shown in Equation 3.2, where K_c and K_i are controller parameters which are changed based on a fixed profile. The controller achieved a decreased offset and increased response time compared to when a constant gain was applied [Hisbullah and Ramachandran, 2002]. However, it also caused oscillatory behaviour at certain times when the tuning was not appropriate.

$$\Delta F_{t+1} = K_c \left(e_t - e_{t-1} + \frac{\Delta t}{K_i} e_t \right) \quad (3.2)$$

An improvement to gain scheduling is on-line adaption of the controller parameters based on available measurement data. The controller is therefore able to adapt to non-predictable system dynamics. In some cases the available measurement data is first processed in order to predict a system parameter or state of interest, which in turn is used in an adaption algorithm. An example of this is hybrid adaptive -Artificial Neural Network (ANN) methods which have been applied to fermentation feed rate control [Duan et al., 2006, Jenzsch et al., 2006b, Jenzsch et al., 2007]. The trained ANN allows estimation of state variables which are otherwise not able to be measured in-line. The ANN output is then used to directly adapt a control law, relating the feed rate to the state variables. As an example, Jenzsch et al. (2006) used a simple control

law as shown in Equation 3.3, where the biomass concentration, X_{total}^* , was found by an ANN trained on 26 batches [Jenzsch et al., 2006b]. In this case, α is the adaptive variable which is solved for by Equation 3.4.

$$F = \frac{\mu_{set} X_{total}^*}{(Y_{xs} - \alpha) S_f} \quad (3.3)$$

$$\alpha = k_1 \Delta X + K_2 \int_{ts}^t \Delta X dt \quad \text{where } -0.15 \leq \alpha \leq 0.15 \quad (3.4)$$

The value of α is used to adapt the yield of biomass on substrate to account for the current state of the system in order to tune the feed rate, F . This is an example of a self-tuning controller, since the state of the system is used to directly calculate the new manipulated variable response.

In the adaptive control methods described, the control law is updated, either based on a pre-determined gain, or based on updates to the process model, where the process model directly influences the control law. In a different approach, it is possible to define the optimal control action which would be desired in order to control a process variable. The control action then aims to minimise the error between this ideal model, and the process outputs. This method refers to model reference adaptive control (MRAC), and has been applied to the problem of feed rate control [Oliveira et al., 2005, Oliveira et al., 2004, Soons et al., 2006]. In this method, the desired response of the system is defined by a model, which would provide the optimal response to input disturbances. This model output, y_m , is then compared to the system output, y , to define an error between the actual response and the desired response. It is this error which is used to tune the controller parameters, by minimising the error term [Landau et al., 2011].

MRAC methods are defined as direct or indirect depending on how the tuning is achieved. In direct adaptive control, the controller variables are adapted based on direct measurements in the system. The error is then calculated directly based on true system variables, and the reference model variables. If the controller parameter values are recalculated based on estimated system parameters, and a model is updated to solve for the desired controller action, then the adaptive control is achieved indirectly. An example of this is described by Oliveira et al. (2004) where a state identification model is used to determine states/outputs of the process from the measured variables. Then the output state is used to define the error from the reference model, and subsequently define the control parameter response [Oliveira et al., 2004].

A challenge for MRAC methods is to determine the Parameter Adaption Algorithm (PAA) to apply in order to adapt the control parameters to achieve not only an optimal response but, importantly, ensure a stable response is given. The original method for this is the MIT rule, which defines how the error changes with the tuning parameters based on the concept of aiming to minimise the squared error [Åström and Wittenmark, 2008]. The MIT rule is an example of a gradient method of adaption. Equation 3.5 shows how the controller parameter θ , is tuned based on the error, e , between the model output and the actual process output. It requires a user defined adaption gain, γ , to be specified, and there is a description of how to determine this value in the literature [Åström and Wittenmark, 2008].

$$\frac{d\theta}{dt} = -\gamma e \frac{de}{dt} \quad (3.5)$$

Another method for adaption is based on stability theory, where methods such as Lyapunov's theory can be used to determine the optimal response of the system and therefore adapt the control response. More details of the MIT method, Lyapunov theory and other PAA methods are given in [Landau et al., 2011, Zeng et al., 1993, Åström and Wittenmark, 2008].

Adaptive control is highly applicable to dynamic systems with high levels of disturbances in

the control parameters. Where other control methods reject disturbance variables, adaptive control inherently adapts for disturbances in the control parameters [Landau et al., 2011]. This is therefore applicable to fermentation systems with unpredictable system dynamics, and unpredictable disturbances. A summary of the adaptive control strategies considered in this review are shown in Table A1.

3.2.3 Model predictive control

Model predictive control (MPC) is a widely used control technique (Forbes et al., 2015), due to its ability to handle complex multivariate systems. Originating in the petrochemical industry, it has now been adopted in a wide range of industries [Qin and Badgwell, 2003], however less so in the biotechnology sector, compared to the chemical industries. MPC requires that a robust predictive process model is in place. The model is simulated up to a time horizon in the future in order to predict the current output, as well as the future development of the system. The prediction is evaluated to define the action required at the current time, based on optimisation of a cost function over the full process time [Stanke and Hitzmann, 2013]. This may be to maximise the production or to minimise the cost, or to follow a trajectory for a certain variable [Seborg et al., 2003]. This optimisation is carried out up to a predetermined point in the future; a defined time-horizon, however only the first step of the optimal solution is implemented. The optimisation is then repeated at the next time interval. In end point MPC, the method is applied to the end point of the batch and then the trajectory along the prediction is monitored on-line [Lauri et al., 2014]. The whole optimisation is also subject to predefined constraints which are built in to the optimisation problem [Seborg et al., 2003].

The key difference between the MPC methods implemented for fermentation applications is the choice of optimisation function. Like traditional control methods, some aim to follow a set point trajectory for a specific variable, such as biomass concentration [Kuprijanov et al., 2013, Zhang and Lennox, 2004] or substrate concentration [Craven et al., 2014].

$$J = \sum_{k=1}^{t+N} (y'_k - w_k)^2 + \gamma \Delta u_k^2 \quad \text{for} \quad t + N < t_{end} \quad (3.6)$$

$$J = \sum_{k=1}^{t_{end}} (y'_k - w_k)^2 + \gamma \Delta u_k^2 \quad \text{for} \quad t + N \geq t_{end} \quad (3.7)$$

Zhang and Lennox describe the application of MPC for a fed-batch penicillin process, in order to follow a predefined biomass trajectory [Zhang and Lennox, 2004]. In Equation 3.6, the objective function, J , is to be minimised, where t is the current sample time, N is the length of the prediction horizon, w_k is the desired biomass concentration at sample time, k , and y'_k is the predicted biomass concentration from the model. Then Δu_k is the change in the manipulated variable, the feed rate, which is made at time k , and γ is a tuning parameter. This equation shows a typical MPC objective function, where the function acts to minimise an error, whilst also achieving the change in a constrained manner, due to the penalisation on large changes in the manipulated variable.

An alternative objective function is to maximise a certain process variable [Santos et al., 2012, Kovárová-Kovar et al., 2000, Chang et al., 2016], rather than follow a trajectory. This is appropriate to situations where the optimal value is not known, but the optimal will be to maximise, subject to operational constraints, which are always an integral part of the MPC problem formulation. A typical example is to maximise product concentration.

Model predictive control is a powerful technique, as decisions are optimal for the full process time, not only at the current time instant, and the impact of disturbances on the system is modelled as part of the optimisation problem. MPC is based on a model, which may be mechanistic or data driven, however part of the success of the method depends on the accuracy of the model, and the ability of the method to deal with disturbances in the system [Laurí et al., 2014]. It is also clear to see that when data driven modelling methods are used, there is a requirement for a higher number of measurement variables [Kovárová-Kovar et al., 2000, Zhang and Lennox, 2004], which is a consideration for the robustness of the method. MPC is considered to be computationally expensive in comparison to other control methods, especially if the optimisations are run for each time point [Laurí et al., 2014]. Although it is a standard method employed in other industries, it seems as though there is a large step change needed if the method is to be applied more widely to biological processes, and work should therefore be invested to produce robust process models. If this is done, it will provide a powerful and flexible control method, which can optimise the process as a whole, and be applied for a range of control objectives. Table A2 provides examples of MPC applied to fermentation systems for feed rate manipulation.

3.2.4 Fuzzy control

Fuzzy control is based on the principles of fuzzy logic, which is designed to deal with uncertainties in systems without utilising complex models. This uncertainty arises in fermentation processes due to the non-linear behaviour of batch and fed-batch processes [Lee et al., 1999]. Quite a different approach to the other model based methods, fuzzy logic requires no initial knowledge of the dynamics of the system. Instead the user's experience with the process is utilised in order to control the process based on an evaluation of the current state of the process.

The principle of fuzzy control is the conversion of quantitative data into qualitative parameters. The procedure is clarified with the following definitions:

Fuzzy set: A linguistic term to define the properties of a numerical variable

Membership function: A value between 0 and 1 which defines the degree to which a certain variable fits a fuzzy set.

The basis of the method is the conversion of numerical data into 'membership functions' based on the degree to which they fit a fuzzy set. For each input variable, a range of values which are possible for that variable are first defined to provide a form of scale to which the data value can be compared. The variable can then be described qualitatively and assigned into a 'fuzzy set' which describes the condition of that variable. As an example, a temperature measurement of 19°C may be compared to a range of 15-27°C and assigned to the fuzzy set 'Cold' with a membership function of 0.7, and 'OK' with a membership function of 0.3. Note that membership functions must always add to 1. The fuzzy set is defined as a matrix consisting of the original value of the variable, and the degree to which it is described in that category, on a scale of 0-1. This assigning of variables into fuzzy sets is based on a defined 'membership function' which is designed to assign each input variable into a set. This process is called fuzzification.

The fuzzy sets are then used to interpret the current state of the system. The basis of fuzzy control is a set of rules based on the condition of the system, which are described from experience with the process. This is in the form of conditional statements using terms such as 'if' and 'then'. From the previous example of temperature, you could define a fuzzy rule such as : If X is 'hot' then Y is 'high' Eg. Where X is the input variable, Temperature, and Y is a flow variable The output of the rules is called defuzzification [Seborg et al., 2003]. This is how the user experience with the process is incorporated into the controller.

Due to the rule base, and the definitions of the fuzzy sets, it can be argued that fuzzy control methods provide some insight into the dynamics of the complex non-linear systems without the need for model identification [Babuška and Verbruggen, 1996]. The outcome of fuzzy control is to identify the current state of the system, and it therefore provides information to the user. This is in contrast to black box models, or artificial neural networks which are specific to certain data they are trained on, and since they provide no real insight into the system, are not adaptable by the user. Fuzzy control can be considered to combine user knowledge and trends, with past data and is therefore somewhat more intuitive to the user. This may be suitable to apply to established biological processes, which have been historically controlled by a more manual approach- where the operators experience with the process has been the primary form of process control [Horiuchi and Hiraga, 1999]. Since the method is based on linguistic rules rather than mathematical systems, they are more simple for users to adapt, which may also make it more applicable to different processes or scales, by minor adaptations to the rule base [Babuška and Verbruggen, 1996]. Despite these benefits, there are limited examples of fuzzy logic applied to the problem of feed rate control, and those references which are available are not the most recent [Hisbullah and Ramachandran, 2003, Horiuchi and Hiraga, 1999, Zhang et al., 1994]. It is clear that there is certainly not a trend towards fuzzy control despite the named benefits. Table A3 provides a summary of the control strategies covered in this section.

3.2.5 Artificial Neural Networks

An Artificial Neural Network (ANN) is a data driven modelling technique which can describe a complex non-linear system without the need for explicit model equations. The method requires past process data, in order to train a network to predict outcomes from inputs to the system. The method is applicable to industrial applications where there is past data, but there may be limited time for model development [Glassey et al., 1994]. It is relatively fast to construct an ANN and it shows good performance for such multivariate problems [Lübbert and Simutis, 1994].

The name is given to this type of control strategy based on the storage of data in nodes and 'neurones' in a similar manner to biological networks. A large input vector of multivariate data is processed via weighted functions to produce a predicted output. The output for each node is defined by the transfer function and a weighting. The functions may, for example, be sigmoidal for smooth output values, or a step function for discrete on/off outputs [Stanke and Hitzmann, 2013]. There are discussions in more detail about types of structures and network algorithms used [Lee et al., 1999]. Based on the selected structure, the network is then trained on process data in order to determine appropriate weighting values for the network. The back propagation algorithm [Rumelhart et al., 1986] is commonly used in ANN training, as shown by [Bošković and Narendra, 1995, Peng et al., 2013].

A trained network describes the output variables, based on a set of input variables. In order to utilise this knowledge for a control strategy it must be possible then to determine an optimal value of input, in order to achieve a desired output. The genetic algorithm [Holland, 1984] can be used to solve this optimisation problem, by optimising the output of the ANN for a given set point, for example a target biomass concentration. This method is applicable to such ANN problems as it is applicable to cases where the objective function is non-linear, which is not suitable for many optimisation algorithms. The method is based on the principle of natural selection.

ANNs have been shown to successfully predict fermentation system behaviour based on measured variables, and this has been applied for control applications. It is possible to apply the ANN as a predictor for a variable of interest, and then incorporate this in to a feedback control

algorithm [Ferreira et al., 2001], however this limits the objective of the control system to set point tracking. An alternative is to utilise the ANN directly in an optimisation algorithm to solve for optimal control solutions [Chen et al., 2004a, Peng et al., 2013]. When an optimisation algorithm is used, it is possible to optimise the system over the full process time, in order to maximise total product or biomass formation [Chen et al., 2004a, Peng et al., 2013]. Table A4 shows examples of feed rate control based on artificial neural networks. A disadvantage of this method in comparison to other control techniques is that the resulting network cannot be interpreted in order to understand relationships between variables, so limited process knowledge is gained [Babuška and Verbruggen, 1996]. It should also be noted that the network is trained on a single scale and operation of the process, and there is limited ability for extrapolation [Lübbert and Simutis, 1994], and it is therefore not scalable, and must be retrained for different scales and processes [Babuška and Verbruggen, 1996].

3.2.6 Probing control

The principle of probing control is to apply perturbations in the manipulated variable, and assess the response in the controlled variable, in order to make decisions of the next set point for the manipulated variable. For instance, when considering feed rate control in fermentation processes, a perturbation in the feed flow rate is applied to the system and the response in the dissolved oxygen signal is analysed to determine the next set point for the feed rate [Åkesson et al., 1999]. There are examples of this method successfully applied on-line for fermentation feed control [Henes and Sonnleitner, 2007, Johnsson et al., 2013, Velut et al., 2007].

For fermentation systems, if an increase in the feed rate causes a drop in the DO concentration, then there is capacity in the system for increased feed. The response of the controller is therefore to increase the feed rate. This process is repeated until there is not a response in the measured variable. The system is then operating at full capacity and feed should be reduced slightly. This strategy is a very interesting approach to the feed control problem, since it is hard to define a set point for the feed which is optimal. This method is considered to be 'self-optimising' in the sense that it is not provided with a set point, but instead finds the optimal point and adapts as the optimal changes. This is interesting when applied to fermentation where a key goal is to feed at the maximal feed rate, however without over feeding and inducing the crab-tree effect [Henes and Sonnleitner, 2007, Johnsson et al., 2013, Dewasme et al., 2011]. It also takes into account disturbances in the system, including disturbances in the feed itself, and adapts to different starting biomass concentrations. It also has the benefit of only requiring a single on-line measurement, DO.

Johnsson et al. (2013) implemented a control law to control the feed based on the frequency content of the DO signal. In this way it is possible to interpret the perturbation in the measured variable into a useful on-line measure of the system (frequency content). A low frequency content means that the system is operating close to its oxidative capacity.

$$F(t) = \frac{K}{h}(C_{set} - C_k) + \frac{h}{T_i} \sum_{i \leq j \leq k} (C_{set} - C_j) \quad (3.8)$$

In Equation 3.8, C refers to the calculated frequency content of the DO response, with sampling rate, h . This equation can be considered in the form of a PI controller, with integral time T_i and gain K . This method also employs gain scheduling as explained in Table A4.

For successful implementation it is important to consider the limitations in oxygen transfer. This method is applied to the exponential growth phase, where the growth is substrate limited. Once the oxygen limitation of the system is reached, this method is no longer applicable,

and then traditional DO control is suitable. A DO control system can be run alongside, or be incorporated into the probing controller strategy [Johnsson et al., 2013].

3.2.7 Statistical process control

The concept of statistical process control (SPC) is to analyse the current process operation based on analysis of past batch performance. The implementation requires no process knowledge, only past batch data from which to form an empirical model [Nomikos and MacGregor, 1994]. It is a monitoring tool which can be applied in order to control the process by identifying deviations from an optimal operation in real time. By identifying correlations between key variables, it is possible to identify early in a process if the operation is deviating from the optimal in a way which is not possible by observation or univariate analysis methods [Albert and Kinley, 2001]. It is also possible to identify if the deviation is likely to cause an impact in final performance, or if it has a low impact. Multivariate analysis techniques are able to identify correlation between variables. Two important methods for analysis of process data are multivariate principle component analysis (PCA) as described in detail in [Nomikos and MacGregor, 1994] and partial least squares (PLS) [Nomikos and MacGregor, 1995]. These multiway techniques are able to be applied to multivariate batch data after decomposition of the three dimensional matrices into two dimensional matrices [Nomikos and MacGregor, 1994].

In PCA a dataset is analysed by transforming the original variables in terms of new principal components, in order to expose relationships between the variables. These new principal components capture the greatest variance in the dataset, such that it is possible to reduce the number of new variables. This shows how the data can be compressed in this method. As an example, applied to fermentation data, Albert and Kinley (2001) showed that PCA based on only 5 variables was sufficient to represent their dataset comprising 17 on-line variables and 53 off-line variables.

In order to evaluate the PCA model, methods can be employed to determine the accuracy of the model. The Hotelling T^2 statistic can be used to measure the variability on the model, and the squared prediction error (SPE) evaluates the error between the model and the data values.

Although there are many literature examples where multivariate methods are applied for process modelling and monitoring [Doan and Srinivasan, 2008, Ferreira et al., 2007, Glassey, 2013, Mears et al., 2016], there are few examples found where statistical methods are directly applied for control [Duran-Villalobos et al., 2016, Albert and Kinley, 2001]. This may be due to some of the disadvantages of the method. For example there is poor extrapolation outside of the conditions used to develop the model, and models must be developed separately for each process and scale.

Despite this, a benefit unique to multivariate analysis is that all variables are monitored, rather than selecting specific variables to control. In this way it allows faults to be detected in all on-line measured variables, which does not just mean deviations from optimal, but includes probe failures or other operational errors, which could otherwise have a large impact on process performance. It can be integrated with a user interface in order to alert operators to deviations from optimal operation in real time. This is valuable for industrial applications, where this could otherwise not be detected until it was too late to return the process to a state where it will achieve the desired titre. For industrial application this method can utilise valuable past data in order to optimise future operations. Table A6 shows the methods discussed.

3.3 Conclusions

Manipulation of the substrate feed rate to a fed-batch fermentation process is an area of focus for process optimisation, as it is an important process input, affecting metabolic rates, and affecting the volume dynamics. It has been shown in the review, that it is a manipulated variable which may be used towards multiple objectives. This review has covered a range of control methods from simple open loop control, to data driven methods, and model predictive control. The aim of the review is to present a summary of the available tools developed for manipulation of the feed rate, and to draw some conclusions on the benefits and disadvantages of such approaches. This may guide future control strategy development. When reviewing the literature surrounding fermentation feed rate control there are many aspects to consider in order to make a constructive conclusion;

- The number of measured variables
- The industrial applicability of the required measurements
- Model requirements
- Historical data requirements
- Tuning requirements
- How generic is the method? Process specific tuning?
- Does the method provide process understanding?
- Intuitive for users?
- Requirement for user intervention
- Development time and cost
- Cost/ benefit

These points may be divided into two categories; variable choice, and method choice. These factors will be discussed below with relation to the works cited.

3.3.1 Discussion of control and manipulated variables

This work is focussed on strategies which may find application in industrial processes, and therefore the number and type of measured variables required is of interest. There are works cited in this review including measured variables which may not be applicable to industrial scales, for example spectroscopy [Craven et al., 2014]. However, this work was applied to mammalian cell culture where the process scale of production is much smaller than industrial biotechnology products, so it could be conceivable to use such a technology in production also. Other control variables include the biomass concentration [Kuprijanov et al., 2013, Zhang and Lennox, 2004] and product concentration [Chang et al., 2016, Kovárová-Kovar et al., 2000] for which there is no reliable on-line probes. It is seen in this review, that ANNs are the most common choice for use as a soft sensor, to obtain on-line values for control variables which are not able to be measured [Duan et al., 2006, Jenzsch et al., 2006b, Jenzsch et al., 2007, Kovárová-Kovar et al., 2000]. This is likely to be due to the speed of model development, however this method is not useful for increasing process understanding. PLS has also been applied for this purpose [Zhang and Lennox, 2004].

The reliable on-line measurements on which to base a control strategy are subject to opinion, but there are some clear generalities which can be made. Dissolved oxygen is a relatively robust measurement which is considered fairly standard in industrial fermentation [Sonnleitner, 2013] and has been utilised widely in the literature. In addition, when we consider industrial scale processes, it is of interest to consider the variables which are not subject to heterogeneity issues, for example, off gas measurements which are not representative of a single point in the vessel. In-line probes provide information about a single location in the system, however there is increasing consideration for the inhomogeneities which can occur at larger scales [Neubauer and Junne, 2010, Wechselberger et al., 2012, Formenti et al., 2014]. For this reason, variables which are derived from on-line gas data are interesting, and they provide important physiological information [Sonnleitner, 2013]. This includes CER, OUR and therefore RQ. Other parameters have also been developed which are able to quantify the metabolic state based on off-gas analysis, namely FQ and OQ [Riisgaard and Andersson, 2013]. FQ is the ratio of carbon moles released as CO_2 to carbon moles added as feed. OQ is the same, but for oxygen moles released compared to carbon moles added as feed.

The feed flow rate was the manipulated variable of focus for this review, however in some multivariate control problems, other variables were controlled in unison. For instance where temperature is also controlled, it may be in order to reduce temperature in order to reduce growth rate, and therefore reduce feed consumption, allowing a reduced feed rate to be applied [Velut et al., 2007]. Where the stirrer speed is also manipulated, this is in order to increase the oxygen transfer rate, which is then balanced with the feed which impacts the oxygen uptake rate, allowing regulation of the dissolved oxygen concentration [Johnsson et al., 2013, Velut et al., 2007].

3.3.2 Discussion of methods

Figure 3.2 shows a general comparison of the control types considered in this review. This is in order to aid control strategy development, by comparing the benefits of the methods, in comparison to their requirements for implementation. Although not an exhaustive list, it may be possible to guide decision making about an appropriate method to consider based on the current competencies regarding available process models, or the amount of past data, for example. For each strategy discussed in this review, the basis of the control strategy is assigned to either historical data, a process model or user experience. Of course this is also open to interpretation, for example all methods require a certain level of experience with the process in order to develop the objectives and the method, however this aims to show the main requirement for each controller type. From the chart, probing control and fuzzy control are highly valuable since the methods do not directly require a developed model, or a vast amount of historical data in order to develop, but is able to deal with unpredictable dynamics and process disturbances. If a process model is available, MPC is desirable as it achieves all the named benefits.

With respect to the benefits of the method, it is case specific how important they are considered to be. Probing control, fuzzy control, model predictive methods, and statistical process control achieve all the mentioned benefits in this summary, since they also provide insight into the process. In industry, this insight into the process is highly valuable, both for future process development, but also for an operator to be able to interpret the control output.

Another aspect to consider, which is not covered in Figure 3.2 is the control objective. From Table A5 it is seen that the objective for the probing control strategy is fixed, as it is integral to the method. However model-based methods may be applied for a wide range of control objectives depending on the process of interest, as can be seen from the range of objectives for the model predictive control applications in Figure A2.

If a process model is available, it is most valuable to utilise this knowledge and develop a control strategy which is flexible, and also may be developed over time as the model develops, such as adaptive control or model predictive control. This is also a very valuable way to consolidate process knowledge. MPC is a method which provides all the named benefits, and also allows for complete freedom in the choice of objective function for the controller. If model-based methods are to be applied more widely in industry, there must continue to be a focus on process model development, and uncertainty analysis, so that there are robust models available which are applicable to control problems in industrial biological processes. This will also require pilot scale testing, of which there is in general limited reference to in the literature.

		Open loop	Probing	Fuzzy control	Gain schedule	Adaptive control	MPC	ANN	SPC
Benefits	Deals with non-linear dynamics								
	Deals with unpredictable dynamics/disturbances								
	Provides process insight to user								
Requirements	Historical data set								
	Process model								
	User experience								

Figure 3.2: Summary of the benefits and requirements of the control strategies covered in this review. For each controller type, one fundamental requirement was selected, to show whether the controller is primarily based on historical data, a process mode, or user experience. Methods are grouped by this requirement.

3.3.3 Considerations for applying a control strategy in industry

For an advanced control system to be applied in industrial situations it must increase the performance such that the cost of implementation is less than the benefit it brings. Bioreactor performance can be quantified by a benefit/cost ratio [Lübbert and Jørgensen, 2001]. It takes time and resources to implement an advanced strategy, especially if models are required. It costs money to buy software licenses, and implement new strategies in existing control software. The benefit must come through an increased profit, for example an increased product yield. Given the physical constraints of the equipment, and the operational constraints of the strain, the aim

is to maximise the productivity [Lübbert and Jørgensen, 2001]. In this work, for example, the objective may be to optimise the manipulation of feed rate in order to increase the product output from each batch.

By tight control of the entire process, the aim of an advanced process control strategy is ideally to reduce the variance in the control variables. By decreasing the variance, the process can be operated nearer to the optimal [Latour et al., 1986]. This can therefore prove a mathematical cost benefit for a given process, and show that just a small increase in average operation by reduced variance, can provide profits. Since this is a trade-off, the cost of implementation should be minimised as much as possible. This means therefore the method should not be so complex, that it requires user intervention, and additional man hours, as this also increases the cost of the control strategy. This should also be a consideration when deciding on a novel control strategy.

Part II

Modelling

4 Data-driven modelling of fermentation processes

In order to gain an understanding of the industrial process of interest in this work, a large historical dataset is analysed. From this, it is possible to gain an initial understanding of the process operation and the control strategies implemented. There is also the possibility to apply data-driven modelling methods in order to gain a more in depth understanding of the process. In this chapter, multivariate analysis is applied to a production scale dataset. This chapter of the thesis is based upon the following article:

Functional unfold principal component regression methodology for analysis of industrial batch process data.

Mears L., Nørregård R., Sin G., Gernaey K. V., Stocks S. M., Albaek M. O., Villez K.
AIChE Journal 62, 6:1986-1994. (2016)

4.1 Introduction

There is increasing interest in the use of big data concepts in order to interpret large datasets and identify trends. Data-driven modelling describes a process where a model is developed by fitting a model structure to available data. Since these methods require no prior knowledge of the system, they may provide a faster approach to model development. This may be especially relevant in an industrial setting, where there is past data available, but there may be limited time available for model development. Although the methods are well established, and applied routinely in other industries, the methods are not yet routinely applied to large scale bioprocesses [Glassey, 2013].

Data-driven modelling covers a wide range of methods which may roughly be divided into two groups; statistical methods (non-linear regression, multivariate analysis) and machine learning methods (artificial neural networks (ANN), support vector machines). Since there is such a large scope to cover, it is decided to focus on multivariate analysis techniques, as they have previously been applied successfully to fermentation process data.

Multivariate analysis methods may be applied on-line as a monitoring tool [Lennox et al., 2000, Zhang and Lennox, 2004, Albert and Kinley, 2001], or off-line in order to obtain process knowledge and identify batch trends [Ferreira et al., 2007, Mercier et al., 2013]. If the model is used as a monitoring tool, this may be in order to identify deviations from a reference batch performance [Lennox et al., 2000, Albert and Kinley, 2001] in order to identify when the process deviates from the optimal path, in a way which is difficult to identify from univariate charts. Alternatively, the monitoring tool may aim to utilise available on-line data in order to predict a variable of interest which is not available on-line in a soft-sensor approach [Zhang and Lennox, 2004]. When applied off-line, to a historical dataset, the aim is to identify trends in the process data which may lead to desirable operating conditions [Ferreira et al., 2007].

4.1.1 Data-driven modelling of batch processes

Modelling of batch processes is of interest for process monitoring and optimisation. A common problem is a lack of available measurements for key process parameters, and modelling tools are therefore desirable to monitor the progress of a batch. Models can also aid process optimisation and may be used in control strategy development. The limitation to model development, especially for batch processes, is a lack of understanding of the system dynamics [Nomikos and MacGregor, 1995]. This can limit the application of mechanistic models, since the non-linear process dynamics are difficult to identify. There is therefore an interest in black-box modelling approaches for batch process data.

Multivariate methods are of growing interest in order to analyse the large datasets which are routinely collected from processing facilities. Multivariate methods such as principal component regression (PCR) and partial least squares (PLS) can be applied to identify trends within a dataset of multiple historical batches in order to provide leads for future process optimisation or as a tool for process modelling [Nomikos and MacGregor, 1994].

Multivariate methods specific for application to batch process data were developed by Nomikos and MacGregor [Nomikos and MacGregor, 1995, Nomikos and MacGregor, 1994]. In this work, a method was developed for structuring the three-dimensional dataset so that it is appropriate for use with the standard multivariate modelling algorithms, such as PLS and PCR. Although multivariate methods are well established for analysis of large datasets, their application to batch processes is less common due to the additional challenges associated with data dimensionality, as well as high measurement noise. There are applications of these methods to batch data [Camacho and Picó, 2006a, Cunha et al., 2002], with a majority of references focussed on online applications, and monitoring [Louwerse and Smilde, 2000, Camacho and Picó, 2006b]. Data mining of such a complex dataset requires additional pre-processing stages, however the importance of these steps prior to modelling is often underappreciated [Gurden et al., 2001].

4.1.2 The challenges of industrial batch process datasets

The challenges surrounding appropriate data handling and pre-processing, are especially relevant to complex industrial data. Industrial scale process data is a combination of on-line and off-line measurements and quality attributes, which are often stored in different databases. In the case of on-line measurements, probes are almost continuously reading new values, however there is a need to limit the volume of data which is actually stored by the data acquisition system. There are therefore data compression algorithms, which are required in order to maintain the maximum information in a data series whilst restricting the data dimensionality [Thornhill et al., 2004]. Different data compression algorithms are implemented in commercial data acquisition software to manage the data that is stored. One such example is that a value is logged only after a specified magnitude of change and in addition after a set amount of time. In the dataset discussed in this chapter, some on-line data series consisted of up to 10,000 data points, whilst others contained less than 200, over the same batch time. This poses a challenge for initial data handling. Datasets can become large, when multiple measurements are logged continuously over a batch running for many days. This means that any proposed data handling tools must consider efficiency as an important factor if it is to become a routine operation in industry.

In addition, variables can be measured with varying degrees of precision, meaning that measurements exhibit different levels of noise. The level of noise captured is also dependent on the data compression algorithm used by the acquisition software [Thornhill et al., 2004]. In the literature there are different methods used to approach this issue ranging from manual outlier

removal [Albert and Kinley, 2001], moving average filters [Le et al., 2012], and data interpolation, showing that there is no consensus on the optimal approach. In many filtering methods used for noise reduction, manual inspection is recommended initially in order to specify a suitable level of filter. This means that a filtering method is applied subjectively, and different users are likely to obtain different results. If a methodology is to be applied in an industrial setting, it is desirable that the tools implemented are flexible and generic across multiple batches and process types, the methods are well documented for consistency, and there are only a few, preferably objective, choices to be made by the modeller.

An additional challenge specific to batch processes, is the occurrence of uneven batch lengths. This may be due to differences between the initial conditions, and process disturbances. It is therefore necessary to decide if the data series should be cut, or alternatively scaled, in order to deal with different end times [Wan et al., 2014]). In this chapter, a methodology is proposed which can be applied when working with large, industrial batch process datasets. Multiple pre-processing methods are discussed in the methods section, which are each applied to an industrial dataset, in order to meet the following challenges:

- Different batch lengths
- Different data dimensionality between batches and between variables in a batch
- Different levels of noise captured between variables
- Appropriate variance scaling prior to multivariate analysis

4.2 Methods

In order to apply multivariate methods, pre-processing is required. The following section describes the methods of pre-processing applied in this chapter.

4.2.1 Time Scaling Methods

The simplest method is to cut all batch data to the length of the shortest batch, however this is only applicable if the batches show only small variations in duration, otherwise key data may be lost. There is also the option to cut the data from the beginning, so that the data which is removed is only lag phase data which contains limited relevant dynamics. If time cutting is applied, then the time index remains in its original scale.

Alternatively there is the option to linearly scale the time, so that the index of a time unit is scaled such that all batches have the same dimensionality, but with different absolute times. This maintains all data, however it affects the interpretation of certain derivative variables such as rates or cumulative values, since time scaling also scales the value of such derived variables.

More advanced methods are available for time scaling or time series alignment, including dynamic time warping [Kassidas et al., 1998, Keogh and Ratanamahatana, 2004], qualitative representation of trends [Villez et al., 2008], and the use of indicator variables [Cinar et al., 2003, Ündey and Cinar, 2002]. These methods are valid extensions to the methodology if time cutting or linear time scaling is not sufficient for a certain application, however they were not applied in this work.

4.2.2 Functional analysis

Functional data analysis (FDA)

Functional data analysis is applied as a method to take a large, unevenly sampled dataset of discrete time points, and produce a smooth functional output. The method is selected as a computationally efficient tool utilising matrix algebra, which is applicable to thousands of data points [Ramsey and Silverman, 2005], and therefore complies with the aim of creating a methodology suitable for industrial datasets. By applying this method it is possible to filter noise from the data, reduce the dimensionality and deal with outliers in a single step [Baert et al., 2012, Chen and Liu, 2001].

In analysis of functional data it is assumed that the data points are discrete and noisy samples of a continuous time series, which can be described by a function, meaning that the data points are interdependent [Ramsey and Silverman, 2005]. The standard model for signal-plus-noise is then applicable, as shown in Equation 4.1, where the data vector, y , is described by a function, x , and subject to noise ε .

$$y = x(t) + \varepsilon \quad (4.1)$$

Functional data analysis, defines this function $x(t)$, using a basis function system, known as a functional basis. A functional basis is a set of basis functions, $\Phi(t)$, which are independent of each other [Hastie et al., 2009]. The function $x(t)$ is defined as a linear combination of these basis functions. In Equation 4.2, K is the number of basis functions, c_i is the i^{th} coefficient, and $\Phi_i(t)$ is the corresponding basis function.

$$x(t) = \sum_{i=1}^K c_i \Phi_i(t) \quad (4.2)$$

The function $x(t)$ gains its complexity from the number and type of basis functions used. The type of basis functions are chosen in order to concisely represent dynamics in the dataset. Fourier basis functions are often chosen to represent periodic data, while spline basis functions are the most common choice for open ended non-periodic data. Using spline basis functions, $x(t)$ becomes a spline function. Spline functions are piece-wise polynomial between specified points usually referred to as knots [Ramsey and Silverman, 2005]. The order of the polynomial and the number and position of knots are the key design parameters for the spline functional basis. This work only considers a spline functional basis for FDA.

The coefficients, c , for the spline function are fitted to the data using an ordinary least squares method with an additional term to penalize roughness in the function output [Ramsey and Silverman, 2005], as shown in Equation 4.3. The penalized sum of squared errors (PENSSE) method is applied since we believe that the time trajectories of the measured variables underlying the noisy measurements are smooth. This property is described mathematically in terms of the second derivative [Ramsey and Silverman, 2005]. This is represented in Equation 4.4 by D^2 , referring to the second derivative of the function, x , of a continuous time series, t . The aim of obtaining a smooth function relates to the objective to represent the process dynamics realistically whilst minimising the effect of noise. Equation 4.3 contains the roughness penalty coefficient, λ , which must be defined for the fitting procedure. The optimal value of λ was determined by means of the leave-one-out procedure discussed in Ramsey and Silverman (2005).

$$PENSSE = (y - \Phi c)'(y - \Phi c) + (\lambda PEN(x)) \quad (4.3)$$

$$PEN = \int (D^2 x(s))^2 ds \quad (4.4)$$

It is important to have a sufficient number of knots in the spline function to ensure that the smoothing effect is due to the roughness penalty and not a result of the least squares fitting. In this work, the knots were evenly spaced, and the number of knots was decided on by inspection of the FDA fit to samples of the dataset. The method of functional data analysis is compatible with the possibility of time scaling, as previously described. If the raw data is analysed by FDA, and an equal number of evenly spaced knots is applied for each batch, then time scaling is automatically implemented. This is shown in Figure 4.1.

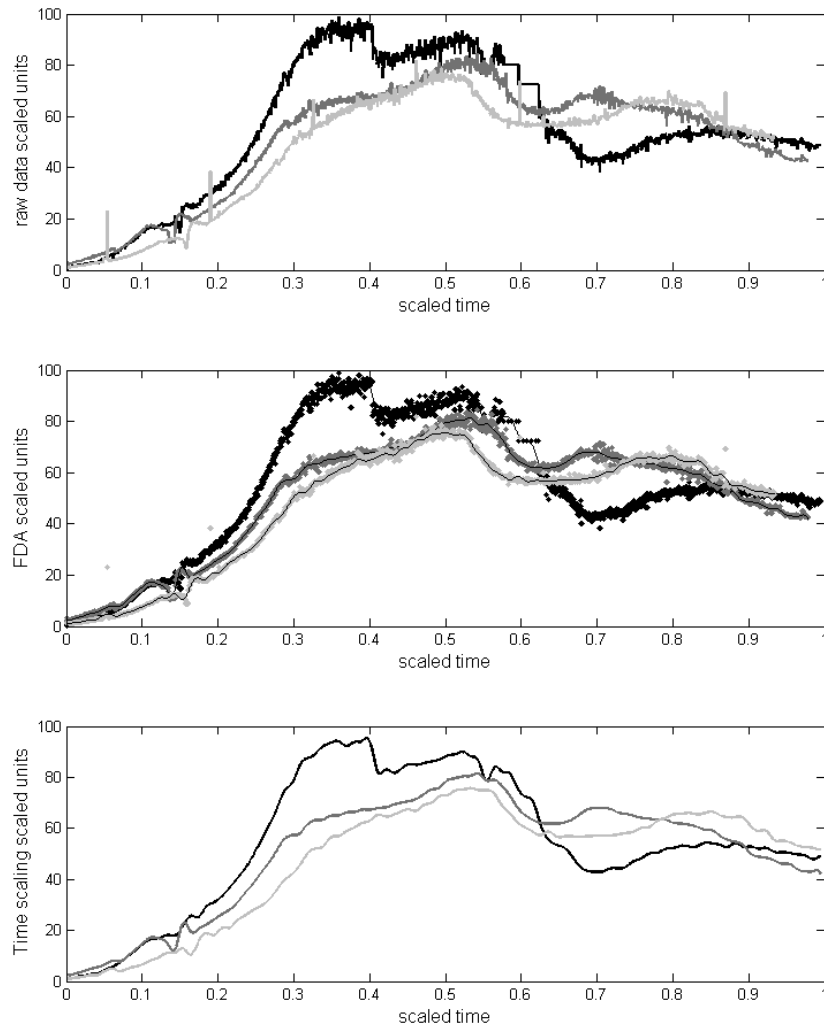


Figure 4.1: Functional analysis, showing a sample of the raw data (top), Functional data analysis result (middle), and the results of time scaling using functional data analysis (bottom). Axis labels are excluded for confidentiality reasons. The different colors show different batches.

Data interpolation

An alternative method of functional analysis is linear interpolation of the data series. Here the data points are linearly interpolated based on a new time series of evenly distributed time points. For comparison, this method is equivalent to a second order spline function with a knot at each data point if the fitting is applied without a roughness penalty. The method of linear interpolation is used as a reference method to assess the success of the FDA procedure using the spline functions presented in the previous section.

4.2.3 Data unfolding

For most multivariate statistical methods, it is required that the data is available as a 2-dimensional matrix. Therefore, given a 3-dimensional matrix with i batches, j time indices, and k variables, as shown in Figure 4.2, dataset unfolding is required [Nomikos and MacGregor, 1995]. Unfolding refers to taking slabs from the dataset to create multiple 2-dimensional matrices of size $[i \times j]$. These are aligned to give a 2-dimensional matrix $[i \times jk]$. This was the only unfolding method applied in this work, since it is the most applicable method for regression of end of batch quality variables [Baert et al., 2012, Glassey, 2013, Mercier et al., 2013].

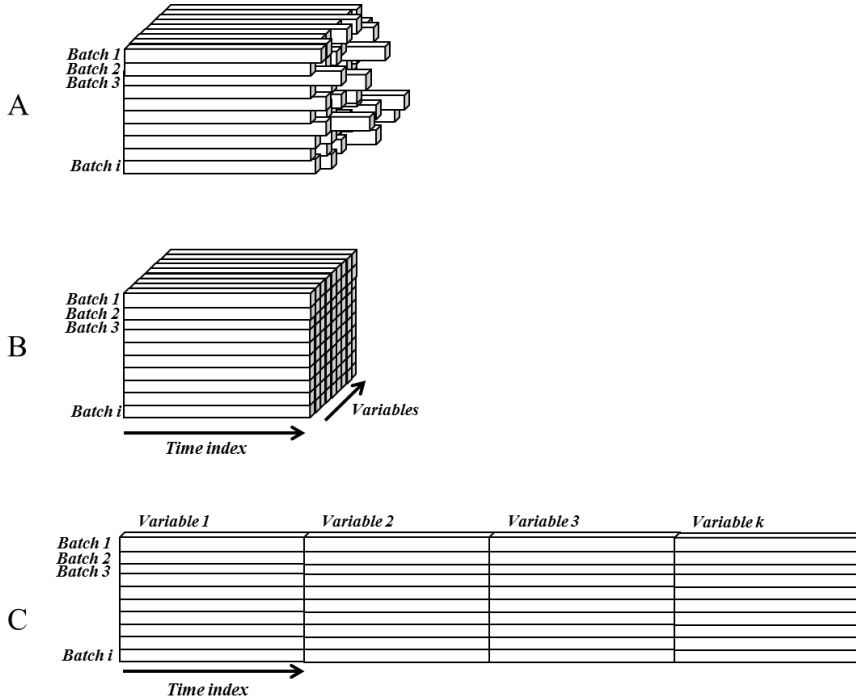


Figure 4.2: Representation of the original data matrix with different sampling frequencies for each batch and each variable (A). The data is then pre-processed by either time cutting and then applying FDA or by time scaling using FDA to create an even and smaller dimension of data for each variable and each batch (B). Unfolding is then implemented (C).

4.2.4 Mean centring

It is necessary to mean centre the data prior to multivariate analysis. In this case, a separate mean is computed and subtracted for each time point within the (warped) batch time span.

4.2.5 Variable Scaling Methods

Commonly used variance scaling methods include column scaling and single-slab scaling (Gurden et al. 2001). Column scaling is applied to each time index, so that each time point has equal variance. In single-slab scaling, all time points for a single variable are scaled for equal cumulative variance. Figure 4.3 shows the effect of the variable scaling method on the data profiles.

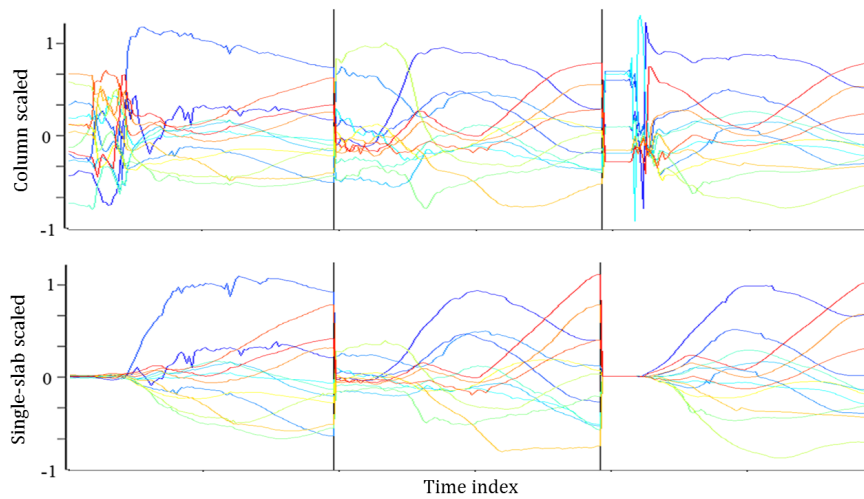


Figure 4.3: Column scaling (top) and single-slab scaling (bottom) applied to three variables which have been mean centred and unfolded. The different colors show different batches. Variable names and time indexes are excluded for confidentiality reasons.

4.2.6 Multivariate analysis

Two commonly used 2-dimensional multivariate analysis methods are PCR and PLS. In this work a comparison is made between the two methods.

Principal component regression (PCR)

The concept of principal component analysis (PCA) is to transform a $[i \times jk]$ data matrix, X , by choice of a new linear basis in the form of principle loadings and scores. The principal loadings constitute the new basis, while the scores represent the data points using this new basis. The principal loadings are defined in order for the first to represent the direction of maximum variance in the data, and the second to represent the direction of second most variance, etc. The dataset is fully represented using $\min(i,jk)$ loadings with corresponding scores. By not including all principal loadings and scores, it is possible to reduce the dimensionality of the data, which then introduces an approximation error. This is represented in Equation 4.5,

where c components are included. Here, the matrix of scores, T , is transformed into the original data matrix, X , by the transpose of the principle loadings matrix, P , with a certain error, E .

$$X = \sum_{i=1}^c t_i p_i^T = TP^T + E \quad (4.5)$$

In this work singular value decomposition was used to obtain the scores and loadings. Principal component regression is an extension to PCA whereby the matrix of scores is used in a regression against a response, y . The regression coefficients, β , corresponding to the scores are obtained by fitting using ordinary least square regression. In Equation 4.6, \hat{y} denotes the estimate of the response.

$$\hat{y} = T\beta^T \quad (4.6)$$

Partial least squares (PLS)

Partial least squares, also known as projection to latent structures, is a similar method of subspace analysis. The fundamental difference between PCR and PLS is that PCR seeks the direction in X which maximises variance, whereas PLS seeks directions which maximize variance and have a high correlation with the response [Hastie et al., 2009]. It therefore identifies variance which is relevant to the regression more directly than PCR methods.

$$X = TP^T + E \quad (4.7)$$

$$Y = UQ^T + F \quad (4.8)$$

$$U(., c) = \beta_c T(., c) \quad (4.9)$$

PLS algorithms are implemented to maximise the covariance between U and T , where U represents the scores for the response, Y . A review of the algorithms can be found in [Andersson, 2009]. In this work the modified kernel algorithm is applied [Dayal and MacGregor, 1997].

4.2.7 Model validation methods

Model validation is required in order to determine the number of principal components (for PCR) or latent variables (for PLS) which should be included in the model. A simple validation method is leave-one-out, whereby every batch is used once as a validation batch. Alternatively, the full dataset can be split into calibration and validation sets, however, the results are then dependent on the division of the dataset. Leave-one-out validation is suitable for regression applications, since it is then possible to see the effect of individual batches on the model prediction at each calibration stage by analysis of the regression coefficients. For this reason, only leave-one-out validation was implemented in this work.

For concluding on the success of a certain model, two metrics are defined. The average error of prediction is calculated as the root mean sum of squared errors (RMSSE) as a percentage of the mean, as shown in Equation 4.10, where \bar{y} denotes the mean value of y , and n is the number of response variables. The standard deviation of the regression coefficients on the mean is also calculated as shown in Equation 4.11, where σ denotes the standard deviation. The average prediction error indicates the accuracy of the model, while the standard deviation of the regression coefficients is a measure of the model robustness. Both values are reported in this paper based on the mean for confidentiality reasons.

$$RMSSE(\%) = \frac{\sqrt{\frac{(y - \hat{y})^T (y - \hat{y})}{n}}}{\bar{y}} \quad (4.10)$$

$$std_{\beta,i} = \frac{\sigma_i}{\beta_i} \text{ for } i = 1, \dots, n \quad (4.11)$$

4.3 Development of a methodology for multivariate analysis of batch process data

A methodology is proposed for analysis of multivariate datasets in order to cover the methods discussed. The input to the methodology is a raw dataset containing multiple time series, of multiple variables, from multiple batches. If time cutting is applied, all data is cut to the length of the shortest batch. If time scaling is applied, the raw data is used for the functional data analysis. Prior to functional data analysis, the order of the function and the locations of the knots must be defined. The data is then smoothed by functional data analysis using a leave-one-out validation procedure. This results in a dataset which has an equal dimension in all variables and for all batches. Unfolding can then be applied, before mean centring and variable scaling. The multivariate method is then applied, using a leave-one-out procedure to determine the number of principal components in the model. The model results are evaluated, and the methodology can be repeated for different choices of pre-processing, and variable choice. Variables are removed if they have low influence on the resulting model, which is shown by a regression coefficient of zero. Figure 4.4 shows the general methodology for applying the methods discussed, such that the methods are implemented in a systematic way.

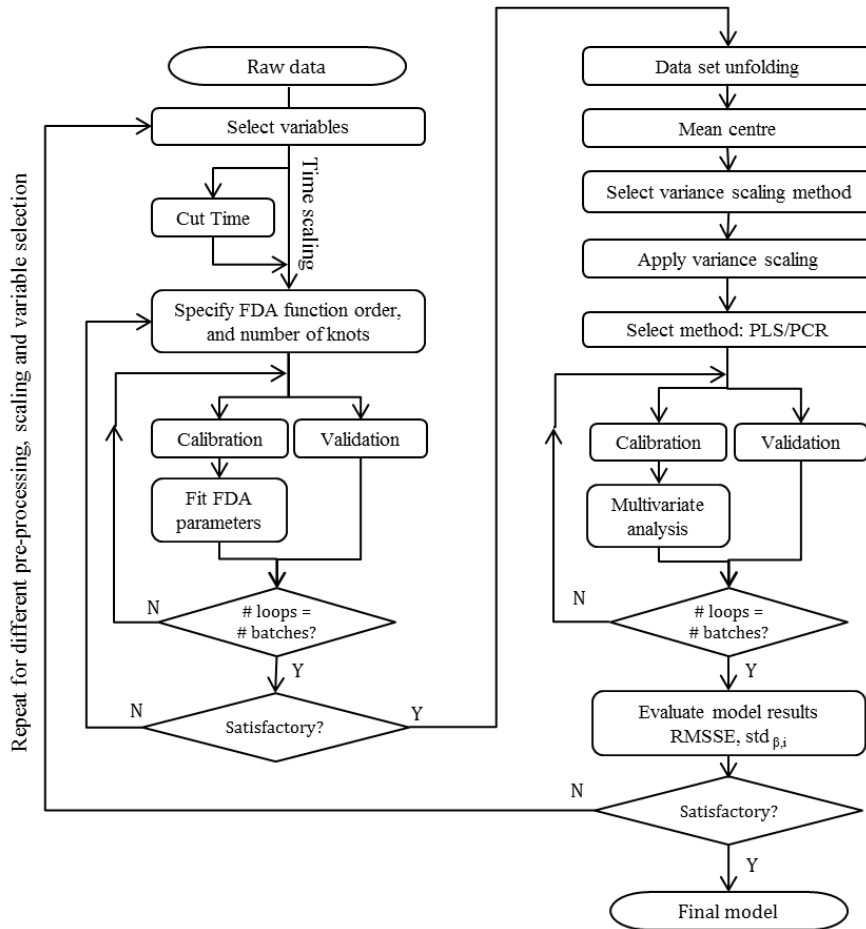


Figure 4.4: Methodology for analysis of batch process data

4.4 Case study: Production data from Novozymes A/S

The case study used in this work, was a production scale fermentation dataset of 30 batches from Novozymes A/S. The dataset was obtained from a fed-batch production process utilising a filamentous fungus strain as the host organism, containing both on-line and off-line measured variables. The aim of the analysis is to predict the final product concentration, which was only measured at the end of the batch.

4.4.1 Results and discussion

The dataset has been analysed by implementing the methodology as shown in Figure 4.4. Multiple iterations were applied in order to implement the different data pre-processing methods, and assess the effect on the resulting regression. The methodology was also applied iteratively for different variable selections. The accuracy of the model was greatly improved by removal of some variables. This shows that certain variables, although showing high variability, did not show variance relevant to the product concentration prediction. The final model included only five on-line variables, making the model interpretation more straightforward for process optimisation.

For the final model, time cutting from the end time to the shortest batch length was found to be most effective. Then FDA was applied, using order 4 spline functions, and applying knots at every 25th time index (no units due to confidentiality). Functions were fitted using a roughness penalty, where the roughness penalty coefficient, λ , was determined by a leave-one-out fitting procedure. After unfolding, mean centring was applied, followed by single-slab scaling, and PCR. This resulted in the regression model shown in Figure 4.5, when each batch is used as the validation batch in a leave-one-out procedure.

By implementing the methodology it has been shown that the choice of applied pre-processing methods has a great effect on the final model result. Starting with this final model, individual elements of the model identification procedure were changed, one at a time, in order to assess the effect of each applied method, with the comparison between methods shown in Table 4.1. In Table 4.1, the result of the final model is compared to the result when one method is changed alone.

Table 4.1 shows that the choice of variance scaling method has the greatest effect on the prediction accuracy, with column scaling increasing both the mean error of the prediction, and the standard deviation of the regression coefficients. Figure 4.6 shows the resulting model fit when only the variance scaling method is changed, in comparison to Figure 4.5 showing the final model regression. This method of scaling is commonly used, however there is the risk that noise is amplified in variables where the overall trend in the data is important but there is high background noise present. For example for a cumulative flow profile, implementing column scaling may amplify noise at the start of the batch, where the variance in the values is lower, and lose important information on the final cumulative flow profile as the variance is scaled equally at each time index. This effect is clearly seen in Figure 4.3 when applied to this dataset. It is likely that this problem is common to many batch process datasets, since the collection of such data is a frequent occurrence.

Table 4.1 also shows the effect of FDA in comparison to linear interpolation. In both cases all other aspects of the modelling work is the same, and since the same interval was used for interpolation and for the positioning of the knots in the FDA method, the results are considered comparable in terms of the information content extracted from the original dataset. Even if a greater dimension of data is included in the interpolation method, it still gives a poor regression

Table 4.1: Average prediction error as a percentage of the mean (%) (Top) and two times the standard deviation of the regression coefficients on the mean (bottom). Final model, refers to the best model found, utilizing time cutting, FDA (order 4 function, knots every 25 time index), single-slab scaling, and PCR. One aspect of the final model was changed at a time in order to assess the effect of the method on the prediction. These changes were; column scaling, interpolation (25 time index), interpolation (5 time index) and PLS. All described time index values have no units due to confidentiality of batch length.

	Final model	PLS	Interpolation		Column scaling
			(5 time interval)	(25 time interval)	
RMSSE (%)	7.4	7.9	10.5	10.9	11.2
$2 \text{ std}_{\beta,i}$	0.39	0.72	0.74	0.80	1.18

performance in comparison to the FDA result, as shown by the result when interpolation is carried out at every 5th time index. This suggests that the smoothing achieved using the FDA method, results in a reliable estimation of the underlying trends in this industrial dataset, by filtering noise in the measurements.

The FDA method is highly suited to process datasets, which comprise multiple variables, each showing different profiles in the data. It allows flexibility in the function output so that the same method can be applied in series to multiple variables, without modifying the procedure. It is able to capture different dynamics in the different variables with a single basis function type defined. It is also compatible in the methodology when considering time scaling. If the raw data with different batch lengths is analysed by FDA with a fixed number of knots per batch, this will automatically implement time scaling. In addition, it is computationally efficient when dealing with highly complex data series, since an analytical solution is available for the spline function coefficients estimates. It is therefore considered a highly valuable addition to the methodology, and an efficient processing tool for industry. It also gives an objective solution to the problem of noise filtering and outlier removal. The limitation is the need to define the number of knots which will be used for fitting, which in this methodology is done by inspection. This may be an area of future work for the methodology, in order to create a completely systematic procedure.

In this work the choice of multivariate method has very little impact on the accuracy of the prediction, however this prediction is achieved with one less principal component for the PLS method, with four principal components compared to five for the PCR algorithm. The lowest absolute error was achieved with the PCR method, as shown in Figure 4.7, and this also gives the lowest standard deviation of the regression coefficients, showing a more robust model.

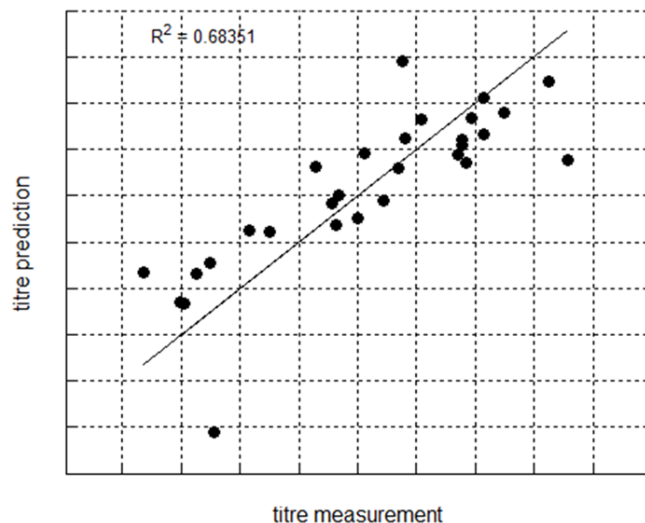


Figure 4.5: Results of the case study, applying the statistical modelling methodology to production scale data from Novozymes A/S. PCR titrer regression model for a 30 batch production scale dataset at Novozymes A/S, using 5 variables. Results show prediction when each batch is used as the validation batch. Time cutting from end of batch time, functional data analysis pretreatment (order 4 function, knots every 25 time index), and single slab scaling. The PCR model uses 5 principal components. Axis labels are excluded for confidentiality reasons.

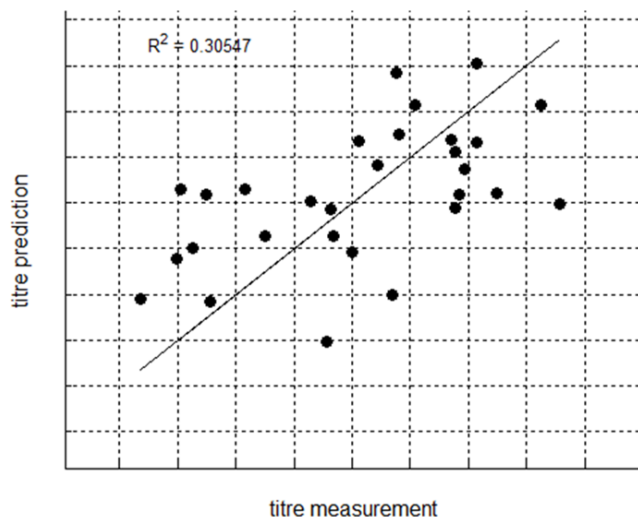


Figure 4.6: PCR titrer regression model for a 30 batch production scale dataset at Novozymes A/S, using 5 variables. Results show prediction when each batch is used as the validation batch. Model is identical to the final model, other than the method used for variance scaling, which in this case was column scaling. (Time cutting from end of batch time, functional data analysis pretreatment (order 4 function, knots every 25 time index), and column scaling). The PCR model uses 5 principal components. Axis labels are excluded for confidentiality reasons.

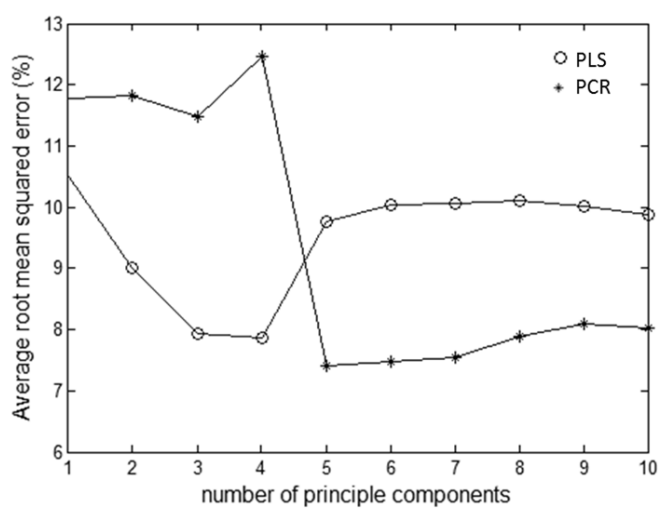


Figure 4.7: Average prediction error as a percentage of the mean (%) against the number of principal components when either PCR or PLS is applied as the multivariate method in the final model, where the model utilizes time cutting from end of batch time, functional data analysis pretreatment (order 4 function, knots every 25 time index), and single slab scaling. Line added to the graph trend for clarity.

Model interpretation

The result of the methodology is a model which can predict the final product concentration with an average error of 7.4%. Once a final model is defined using the methodology, it is then possible to analyse the regression coefficients for the prediction, in order to understand the trends which result in a higher prediction. Figure 4.8 shows the regression coefficients for this model, which correspond to the five variables. It is especially valuable, that the results show time dependent trends, such that desirable conditions over time are identified. Although the details of this optimisation procedure are not discussed for confidentiality, the concept of applying such tools for the purpose of process modelling and optimisation is shown. For example, understanding the cause for the high impact of Variable 1 on the prediction near the end of the batch is shown to be a lead for process optimisation. The end conditions for Variable 2 and 3 are also important, with a higher than average value for Variable 2, and a lower than average value for Variable 3, giving a greater prediction. By analysing the average batch trend in these variables, and identifying which operating conditions lead to this result, future batches can be optimised based on this knowledge.

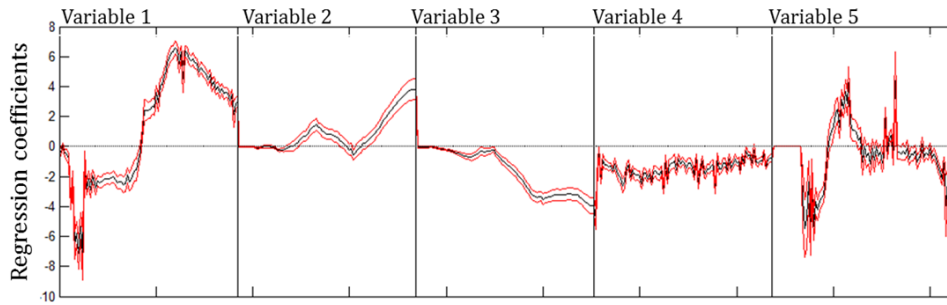


Figure 4.8: Regression coefficients for the 5 variables used in the final model against time index. The black line shows the average regression coefficient, and the red shows the standard deviation as each batch is used as a validation batch. The larger the regression coefficient, the larger impact on the regression. Zero shows that the value of the variable at that time has no impact on the prediction.

Since the goal of the methodology is process optimisation, the predictive capability of the model is of vital importance for the applicability. It is therefore interesting to assess the accuracy of the model, if the two highest titre batches are excluded from the dataset for model building, and the resulting model is applied to these batches. This has been implemented, with the results shown in Figure 4.9. The prediction accuracy is within the range of the calibration batches, which shows that the model is robust to this type of optimisation procedure.

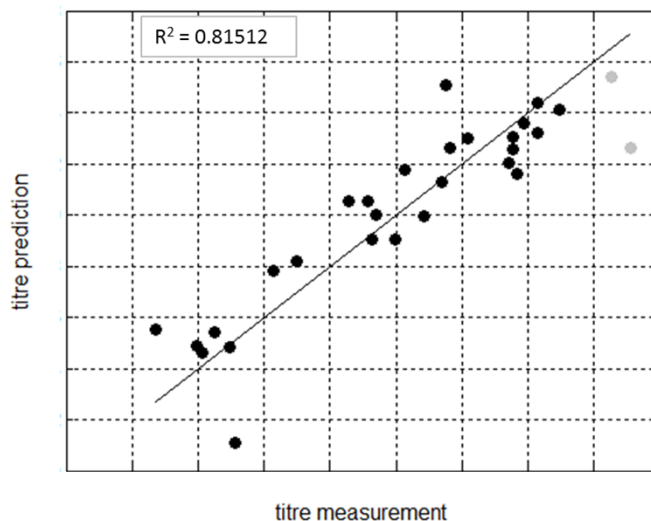


Figure 4.9: PCR titer regression model for a 30 batch production scale dataset at Novozymes A/S, using 5 variables. Black data points used also for model calibration, grey data points only used for model validation. Time cutting from end of batch time, functional data analysis pretreatment (order 4 function, knots every 25 time index), and single slab scaling. The PCR model uses 5 principal components. Axis labels are excluded for confidentiality reasons.

4.5 Conclusions and future work

Multivariate analysis tools were successfully applied in order to model a production scale dataset, and a methodology is developed due to the practical challenges faced when applying conventional multivariate analysis tools to an industrial dataset. The aim of the methodology is to provide a new framework for approaching multivariate analysis of batch process data. A structured approach becomes increasingly important due to the wide range of alternative methods for both multivariate analysis and pre-processing. The methodology is then valuable, in order to investigate a range of these methods for each stage of the pre-processing, and provide an objective solution to noise reduction. The methodology also provides an outline for documentation of the resulting model. This is considered highly valuable for application in an industrial setting.

This chapter discusses how pre-processing is an important and integral stage in multivariate analysis of batch process data. The choice of methods has an effect on the accuracy of the resulting multivariate model, which is investigated using an industrial fermentation dataset from Novozymes A/S. The final product concentration is predicted from a production scale dataset with an average prediction error of 7.4%. The success of the model is dependent on the data pre-processing applied, more than the choice of multivariate method. This result may be especially relevant to those using commercial modelling tools, as it is important that users are aware of the pre-processing applied in the software, and consider the implications. This is especially true of the variance scaling method, as it is shown to have a large effect on the resulting model.

The focus for future developments is now on data management and accessibility, to allow such tools to become applied more widely in industry for process optimisation. It can take a significant amount of time to access the data, both on-line and off-line. Such tools may be more

readily applied if more focus was put also on database management and coordination of data from different sources.

In this work, the modelling tool was applied for the purpose of identifying possible optimisation leads, and for this purpose it has been successful. As discussed in Chapter 3, statistical modelling methods may also be applied for the purpose of control. For control applications, the purpose is usually to follow a fixed batch trajectory of the best performing batch. This may then be applicable to a process, for example producing a pharmaceutical product, where the process operation is very fixed, and there are strict requirements for control of the process to meet product quality requirements. In the field of industrial biotechnology, the drivers for the control strategy are not the same in this way; it is possible to continually improve a process also in the production scale. It is therefore not ideal to consider statistical process control methods in this work. However, it has been shown that statistical modelling tools are highly valuable for the purpose of gaining valuable process insights from the vast amount of process data which is available in an industrial setting.

5 Mechanistic modelling of fermentation processes

The previous chapter has described the development of a data-driven model for the industrial batch process. This has been useful in order to gain process insights, but it is not the ideal platform for control strategy development. In this chapter, a mechanistic model approach is investigated. This is based upon a model which was developed by Albaek et al. (2011) and Albaek et al. (2012). In this chapter the model is described, and uncertainty and sensitivity analysis is conducted.

5.1 Introduction

Mechanistic models provide a mathematical description of the current understanding of a dynamic system, and are therefore important tools for process development [Fernandes et al., 2013]. By constructing a mechanistic model it can aid with consolidating the current process understanding, and they may be developed over time as additional process insights are obtained. There is increased interest in mechanistic modelling due to the Process Analytical Technology (PAT) guidance and the principles of Quality by Design (QbD) which are of focus in the pharmaceutical industry [Gernaey et al., 2010]. Mechanistic models are therefore powerful in this industry due to the information they can provide.

A mechanistic model is formulated based on first principles, where the system is described by a series of equations which are defined deterministically, and have a physical meaning. For bioprocesses, this requires an understanding of the metabolic processes which underpin the system, although the model equations are often heavily simplified. The model can also be extended to include physical relationships for oxygen transfer, heat transfer and mixing. The model should capture the dynamics of the system, however if it is too complex the number of variables to determine becomes infeasible, and can lead to parameter estimation issues.

Mechanistic models for fermentation processes can be described as structured/unstructured and segregated/unsegregated based on the treatment of the cell population in the model [Gernaey et al., 2010]. If the cells are treated as a multi-component system, then the model is described as structured. Although much more simplified, unstructured models are commonly used. If the cell population is treated as a single entity then the model is described as unsegregated, and is therefore more simple in construction. In segregated models, the individual cells are considered heterogeneous. This adds vast complexity to the modelling problem.

When evaluating a mechanistic model it is important to carry out model uncertainty and sensitivity analysis, in order to determine the uncertainty of the parameters [Fernandes et al., 2013]. This can also provide insight into model parameter correlations, and also indicate if the model structure is suitable. Mechanistic modelling can be used to investigate a design space for a system of interest, however it is vital to conduct a thorough analysis of the limitations of the model.

5.1.1 Application of mechanistic models to fermentation processes

Mechanistic models require a significant investment of time, and resources, however their application to multiple stages of the process development and operation can make this investment highly valuable. It is possible to re-use an established mechanistic model to different stages of fermentation process development; planning, process design, monitoring and control. For example, a single mechanistic model structure can be adapted for the following applications:

Off-line applications:

- Definition of process conditions and sensitivity analysis
- Definition of batch initial conditions and start fill
- Off-line control strategy testing

On-line applications:

- On-line state estimation
- On-line control

Although there is a longer development time for such modelling methods in comparison to black-box techniques, the wide range of applications makes them a highly valuable tool for fermentation research and development. There may also be additional benefits, for example Albaek et al. (2012) showed that mechanistic models may be relatively simple to adapt to different host organisms by changing the model parameters [Albaek et al., 2012, Albaek et al., 2011]. In addition, a range of different control strategies may be tested for feasibility prior to on-line testing, including multivariable control. Finally, in a research environment, where collaboration is important, developing mechanistic models provides a platform for knowledge sharing and consolidation of existing understanding.

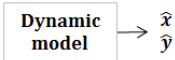
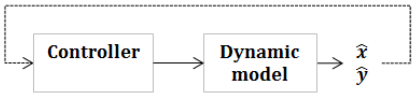


Structure	Applications
	<ul style="list-style-type: none"> • Planning start fill • Planning process conditions • Modelling process sensitivity to process conditions • Analysing process dynamics
	<ul style="list-style-type: none"> • Testing and tuning controller configurations • Designing new process control strategies
	<ul style="list-style-type: none"> • On-line monitoring and state estimation
	<ul style="list-style-type: none"> • On-line control based on state estimator prediction • Predictive control utilising dynamic model

Figure 5.1: Summary of possible mechanistic model structures and their applications. U indicates input variables.

Off-line applications

Definition of process conditions Planning of processes operated at different locations is an industrial challenge, since it is desirable to create a reproducible process operation, despite potential differences in equipment. The different equipment may also have different physical operating limits, for example headspace pressure, aeration rates or stirrer speeds, which are known to be important parameters for fermentation processes [Albaek et al., 2011]. It is important to account for these differences in the batch planning stage, in order to allow for processes which are as reproducible as possible. In order to properly design scale down studies, it is also important to have an understanding of equipment limitations at different scales of operation [Stocks, 2013]. Models are applicable at this stage to assess the process sensitivity to changes in the process conditions, and also assess the effect of different reactor dimensions or technologies, for example impeller types [Albaek et al., 2011]. It is possible, for example, to assess the change in heat and mass transfer rates in different tanks, under different process operating conditions.

Definition of batch initial conditions Industrial fermentation systems are often operated in fed-batch mode which poses specific challenges for planning and scheduling. The batch phase of the process may vary considerably in length due to variations in the lag time, as well as other inherent batch-to-batch variation in initial conditions [Cinar et al., 2003]. Deviations in batch duration, and also variations in final batch fill, lead to variation in product mass achieved in each batch operation. It is desirable to always achieve full vessel capacity to maximise the productive volume of each tank in every fermentation run, within a given process time. A mechanistic model may be applied in order to predict a suitable start fill for a process. Other relevant initial conditions include substrate concentration and biomass concentration. The effect of these initial conditions may also be investigated by means of a mechanistic process model.

Off-line control strategy testing A major benefit of a mechanistic model over a data driven model is that it is dynamic and predictive, and that it better extrapolates outside of the conditions used to develop the model [von Stosch et al., 2014]. For these reasons, mechanistic models are applicable for testing of different control strategies. By addition of a control algorithm coupled to the dynamic process model, it is possible to simulate different controller algorithms prior to on-line implementation [Birol et al., 2002]. Typical controlled variables for a fermentation process include the temperature, pH and dissolved oxygen concentration [De Leon et al., 2001, Alford, 2006]. These are typically controlled using PID controllers. For example, the dissolved oxygen concentration may be controlled using the feed rate as manipulated variable [Bodizs et al., 2007, Lee et al., 1999, Albaek et al., 2011], or the air flowrate as the manipulated variable [Gnoth et al., 2008, Gomes and Menawat, 2000]. The control algorithm for off-line implementation would then be a simple PID control implementation utilising the dynamic model output of the DO vs a reference DO set point trajectory in order to define the manipulated variable in a closed loop. By implementing this off-line, it allows tuning to be simulated rather than utilising resources with on-line testing and tuning. Although fine tuning may still be required on-line, this will greatly reduce the cost and time of control strategy tuning.

On-line applications

On-line monitoring There is a lack of robust, on-line sensors for key parameters of interest in the field, such as substrate, product and biomass concentration [Sonnleitner, 2013], due to challenges specific to the development of in-line sensors which are applicable to industrial fermentation systems. There is therefore an interest in state estimators which utilise reliable and available on-line measured variables to predict the unknown states in real-time [Sagmeister et al., 2013, Luttmann et al., 2012]. Mechanistic models provide a basis for development of on-line monitoring tools.

On-line control Mechanistic process models may be applied on-line for the purpose of control. These model-based control strategies may be beneficial due to the multivariate nature of the control problem. By decoupling the control problem into multiple closed loops it does not consider the full control objective as can be done with a model-based approach. There are also a wide range of control objectives which can be met by application of a model-based control strategy, as discussed in Chapter 3.

Figure 5.1 provides a summary of the applications described, and the model structure employed in each case. Due to the wide range of applications, mechanistic model development may be highly valuable in an industrial environment, as they may be applied for a range of applications.

5.2 Methods

5.2.1 Model description

A mechanistic model has been developed to describe industrial fermentation processes operated at the pilot scale facility at Novozymes A/S [Albaek et al., 2011]. This model was originally developed to describe an *Aspergillus oryzae* process [Albaek et al., 2011], and was then adapted in order to describe a *Trichoderma reesei* process [Albaek et al., 2012]. The following section describes the model, however for full details the publications by Albaek et al. may be consulted. The model is implemented in MATLAB [The MathWorks Inc., 2013] and solved using the built in solver, ode23s.

The model of interest is from the following publication:

Evaluation of the energy efficiency of enzyme fermentation by mechanistic modeling.
Albaek M. O., Gernaey K. V., Hansen M. S., Stocks S. M.
Biotechnology and Bioengineering 109:950-61. 2012.

Volume balance

The total liquid volume in the system is described by the following ordinary differential equation, taking into account mass loss due to evaporation and dissolution and evolution of gases from the liquid phase.

$$\frac{dV}{dt} = \frac{\rho_F F - 1000 F_{evap} + \rho_F Y_{SO} F - \rho_F Y_{SC} F}{\rho} \quad [L/h] \quad (5.1)$$

Feed rate

The feed rate is applied to the system using a PI controller, where the dissolved oxygen is the measured variable, and the feed rate is the manipulated variable. This results in an additional state in the model for the integral term, I, which is cumulative over the process time. The feed is subject to maximum and minimum feed rates, which is the same as in the process operation.

$$F = F_0 + K_c * (DO_{set} - DO_{measured}) + I \quad [L/h] \quad (5.2)$$

Biomass concentration

The biomass concentration is modelled using a growth rate term. The growth rate is calculated in an unconventional way, by assuming no accumulation of substrate in the system. In this way

the growth rate is solved by assuming all feed is consumed for either biomass accumulation or maintenance. This assumption is made in relation to the feed rate control, which is added based on a dissolved oxygen profile. The assumption is then that the feed is consumed immediately as it is added, and is as such the limiting rate factor.

$$\frac{dX_v}{dt} = \mu X_v - X_v \frac{dV}{V} \quad [g/L.h] \quad (5.3)$$

$$\mu = \frac{\frac{S_F F}{V} - S_v \frac{dV}{V} - m_s}{\frac{X_v}{y_{xs}}} \quad [h^{-1}] \quad (5.4)$$

Viscosity

Viscosity is calculated in the same manner as discussed in Albaek et al. 2012 [Albaek et al., 2012], where the parameter fitting is also described. The viscosity is a function of the biomass concentration, and the stirrer speed, which is important in this non-newtonian fluid. k_s is the shear rate constant [Metzner and Otto, 1957], which is defined as 11.

$$\mu_{app} = C_1 X_v^{a_1} (k_s N)^{C_2 X_v^{\beta_1} - 1} \quad [Pa.s] \quad (5.5)$$

Mass transfer coefficient

The mass transfer coefficient, $k_L a$, is calculated in the same manner as discussed in Albaek et al. 2011, where the parameter fitting is also described. The mass transfer coefficient is a function of the apparent viscosity, as described in equation 5.5. It is also a function of the power input to the broth per liquid volume, as well as the superficial gas velocity at actual temperature and pressure, v_g .

$$k_L a = C \frac{P_{broth}^{0.52}}{V} v_g^{0.15} \mu_{app}^{-0.50} \quad [h^{-1}] \quad (5.6)$$

The power input per unit volume is required for this empirical formula. This is also a term which is commonly applied to define the mixing in a fermentation process. Power consumption from agitation is given by the following power number relation. In this relation, the unaerated impeller power number is defined by Albaek et al. 2011. The term P_g/P_o defines the gassed power input in relation to the ungassed power input, and this is determined to be 0.8 [Albaek et al., 2008]. The density of the broth is a constant, ρ , and the number of impellers is given by n .

$$P_{agitation} = 0.001 n P_o \rho N^3 D^5 (P_g/P_o) \quad [kW] \quad (5.7)$$

In addition, power is dissipated by aeration [Roels and Heijnen, 1980]. The power dissipated by aeration, P_{air} is calculated as discussed in Roels and Heijnen (1980). In this equation, vgn , is the superficial gas velocity at normal temperature and pressure, g is gravitational force, and Z is the ungassed liquid height. The scaling factor of 22.4 (L/mol) refers to the volume of one mole of gas at standard temperature and pressure.

$$P_{air} = 0.001 \frac{V v g n R T}{22.4 Z} \ln \left(1 + \frac{\rho g Z}{p_{head}} \right) \quad [kW] \quad (5.8)$$

The value of P_{broth} is then the sum of the power input by aeration and the power input from mixing.

$$P_{broth} = P_{air} + P_{agitation} \quad [kW] \quad (5.9)$$

Dissolved oxygen concentration

The dissolved oxygen is calculated based on a mass balance between oxygen in the inlet gas which dissolves into the liquid phase, and the oxygen consumption from the liquid phase due to respiration. This requires knowledge of the driving force for oxygen transfer into the liquid phase. This is approximated based on a logarithmic mean basis. In Equation 5.10, DO^* represents the saturation concentration of oxygen, and this is given at the inlet and outlet conditions, based on the pressure and the oxygen content of the air.

$$\frac{dDO_v}{dt} = k_L a \frac{(DO_{v,in}^* - DO_v) - (DO_{v,out}^* - DO_v)}{\log \frac{(DO_{v,in}^* - DO_v)}{(DO_{v,out}^* - DO_v)}} - (\mu y_{xo} + m_o) X_v - \frac{DO_v}{V} \frac{dV}{dt} \quad [molO_2/L.h] \quad (5.10)$$

Product concentration

The product concentration is calculated based on the biomass growth and the yield of product from substrate. In this way the product formation is assumed to be only growth associated.

$$\frac{dP}{dt} = \mu X_v \frac{1}{Y_{SX}} Y_{SP} - P_v \frac{dV}{V} \quad [g/L.h] \quad (5.11)$$

5.2.2 Uncertainty analysis

In order to critically evaluate a model, and identify the robustness of a model, it is important to conduct uncertainty analysis. Since the model is defined with a number of parameters which are then fitted to experimental data, it is important to identify the accuracy of the model structure and the uncertainty in the parameters. In the work of Albaek (2012), expert review in combination with Monte Carlo simulations were used to assess the model uncertainty. In this work, this method is compared to the bootstrap method for uncertainty analysis. This provides two different perspectives on the expected uncertainty in the model.

Expert review process

A commonly used method for parameter uncertainty analysis is expert review, whereby the uncertainty is estimated subjectively [Sin et al., 2009]. This is the method used in the case study in order to define which parameters were expected to show low, medium or high levels of uncertainty, corresponding to a variation of 10%, 30% and 50% respectively. This method is the most simple to implement for a model with many parameters, but, apart from available process knowledge, there is limited justification for the level assigned, and there is also an assumption made about parameter distribution. This method assumes a normal distribution, and no correlation between the parameters. In many cases this is likely to be an oversimplification. The expert review process is followed by Monte Carlo simulations in order to assess the model uncertainty due to the parameter uncertainty levels defined.

Bootstrap sampling

The bootstrap method [Efron, 1979] is a method of parameter estimation including parameter estimation uncertainty. The data points are assumed to be function of a set of true parameters, and noise, as described in Equation 5.12. The model output, $f(\theta)$, is first fitted to the dataset (y), by minimising the sum of squared errors when normalised by the mean of the data to account for the different scales, as shown in Equation 5.13. The residuals, ε , between the model

output and the data are identified.

$$y_i = f_i(\theta) + \varepsilon_i \quad (5.12)$$

$$\hat{\theta}: \text{minimize } \left\| \frac{f(\theta) - y}{\bar{y}} \right\|^2 \quad (5.13)$$

$$\varepsilon = f(\theta) - y \quad (5.14)$$

In the method, the vector of residuals, ε , between the model fit and the data are sampled, by a method of random sampling with replacement. This re-sampling creates a new dataset for fitting, and the model parameters are fitted again. 100 iterations of re-sampling with replacement were carried out. Each new dataset which is generated provides a new parameter set, θ_i , such that the final mean, standard deviation and correlation of the parameters is defined. This method assumes that the residuals between the model prediction and the experimental data is due to measurement errors, therefore it is only applied to variables which this is applicable for. Bootstrap sampling is applied to the biomass, product, viscosity, mass transfer coefficient and the weight data. Growth rate is not used, because it is already captured indirectly in the biomass data, and feed and dissolved oxygen are not used as these are control parameters and this assumption does not hold for our model prediction. The bootstrap method is applied as a comparison to the expert review process.

Monte Carlo simulations

The expert review process and the bootstrap method both provide an estimate of the parameter uncertainty. In order to assess the impact of the parameter uncertainty on the model output, Monte Carlo simulations are used to simulate the model for the parameter space defined. The parameter space is found by application of Latin Hypercube sampling [Helton and Davis, 2003] where 100 samples are selected. For the expert review process it is assumed that there is no correlation between the parameters, and for the bootstrap method the correlation matrix obtained from the parameter fitting is applied. The Monte Carlo simulations then show the model output for the parameter space defined.

5.3 Results and discussion

Uncertainty analysis was carried out on the model described in the methods section. A summary of the model structure is provided in Figure 5.2. The model has ten parameters, five state variables and eight outputs. As previously discussed, in the original model implementation, an expert review process was carried out which resulted in the parameter uncertainties, as shown in Figure 5.2.

5.3.1 Expert review

Firstly, the uncertainty analysis was carried out using the expert review uncertainties. This has already been completed in the work of Albaek et al. (2012), but by repeating this, it ensures that the model implementation is correct in this work, and the results are reproducible. The results of the 100 Monte Carlo simulations, as shown in Figure 5.6, are the same as shown in the work of Albaek et al. (2012). There is a relatively high uncertainty in the model output. This is especially prominent in the product concentration and the biomass concentration. The general trend captured in the model is in agreement with the experimental data, however there

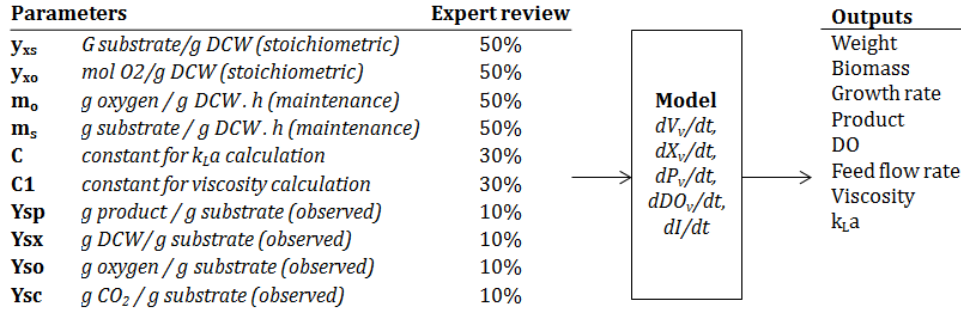


Figure 5.2: Albaek et al. 2012 model representation, showing the model parameters (left), model states (centre box) and outputs (right). The parameters described as (observed) are estimated from experimental data by a regression fit. An expert review is conducted for the parameter estimation in the work by Albaek (2012), and these values are also provided next to the parameters

is relatively high uncertainty.

The parameter uncertainty estimates defined in the expert review process were in some cases data-based, for example the observed yields. This means that the uncertainty defined by the expert review process covers both model parameter uncertainty and data uncertainty and batch-to-batch variations in the parameters. In this way, a large parameter uncertainty estimate is relevant, as it represents the parameter uncertainty when applied to multiple batches and at different conditions. The large model output uncertainty is not only due to the magnitude of the parameter uncertainties, but also due to the fact that the correlation between variables is neglected. This means that a wider parameter space is sampled than is relevant for the model uncertainty analysis.

5.3.2 Bootstrap method

The uncertainty analysis was then repeated, but this time using the bootstrap method to obtain the model parameter uncertainties. Figure 5.3 shows the results of the bootstrap sampling, where 100 iterations of parameter fitting were completed. The grey lines show the residual re-sampling, and the red line shows the average model fit obtained when re-sampling. Firstly, it is noted that the average parameter values obtained from the bootstrap method provide a better overall fit to the data than the values defined from the original expert review process. This can be seen by comparing the average fit obtained from the original Monte Carlo method in the top figure of Figure 5.6 to the average fit of the bottom figure, where the average parameter values are applied to the model. This is especially true of the biomass and product concentration profiles.

The exact parameter values can not be disclosed for confidentiality reasons, however it is important to identify if the new set of fitted parameters lies within the expected range for the parameters, to ensure that they still have the same physical meaning in the model. This can be shown by comparing the original parameter values, with the average value obtained from the bootstrap method. The average parameter values obtained by fitting were all within the ranges defined in the original expert review process, other than Y_{sc} which was just 7% greater than the upper limit from the expert review process. The fact that the average values are within the expected parameter ranges shows that the relevance of the parameters is maintained, and the

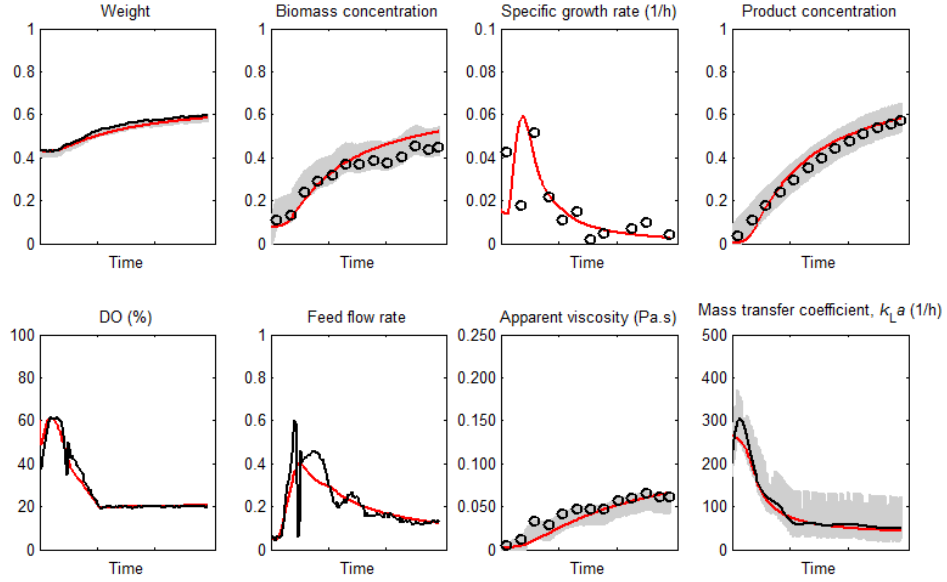


Figure 5.3: 100 bootstrap samples of the model residuals (grey) and the model output using the average parameter values (red) showing that the objective function for parameter fitting was successful.

values of the parameters have physical meaning in the mechanistic model.

The distribution of the parameter estimates indicate the uncertainty on the parameter estimates. Figure 5.4 shows the distribution profiles for the ten parameters. The percentages show two times the standard deviations as a percentage of the mean value. This can therefore be directly compared to the variation range applied to the Monte Carlo simulations in the benchmark uncertainty analysis. A benefit of applying the bootstrap method is that the correlation between the coefficients can also be analysed. This information is provided in Figure 5.5. A summary of the parameter uncertainty estimates obtained from expert review, and from bootstrap sampling is given in Figure 5.7.

The observed yield parameters, Y_{sx} , Y_{sp} , Y_{so} and Y_{sc} , were estimated to be in the low uncertainty class in the original expert review process, meaning an expected uncertainty range of 10%. Figure 5.7 shows that in all cases the bootstrap uncertainty is over double this value, and for Y_{so} the uncertainty reaches 77%. The parameters Y_{so} and Y_{sc} are applied in the volume balance, in order to account for dissolution and evolution of gases into the liquid phase. It is likely to be that the high uncertainty in these two parameters is due to the trade-off between the two parameters; an increase in one may be balanced by a decrease in the other. This can also be seen in the negative correlation coefficient of -0.3 in Figure 5.5.

The observed yields for the biomass and product on substrate also have higher uncertainty levels than expected in the expert review process, but more importantly it is observed that these parameters have a correlation coefficient of 0.97. Both parameters are only involved in the calculation of the product concentration. By analysis of Equation 5.11 it is clear that this correlation is due to the equation structure, where the product concentration is proportional to

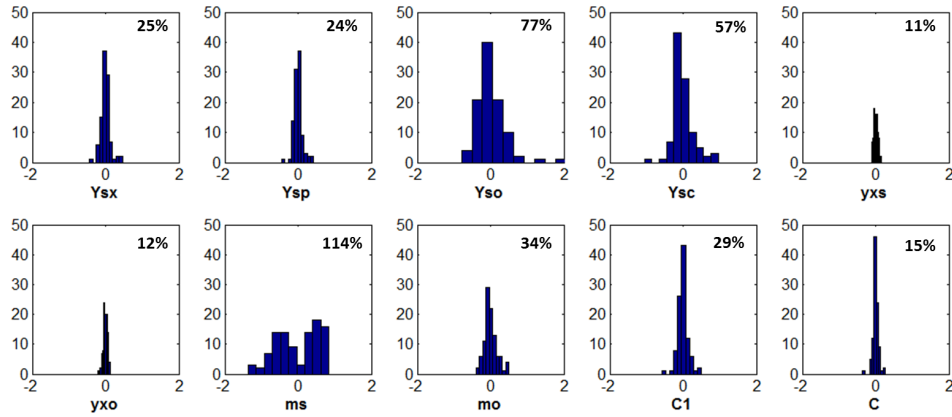


Figure 5.4: Parameter distributions after 100 bootstrap samples shown relative to the mean value. 2 standard deviations as a percentage of mean is shown for each parameter.

	Ysx	Ysp	Yso	Ysc	yx	yo	ms	mo	C1	C
Ysx	1	0.97	-0.19	0.07	-0.01	-0.03	-0.17	-0.23	-0.43	-0.42
Ysp		1	-0.22	0.07	0.02	-0.01	-0.08	-0.08	-0.46	-0.48
Yso			1	-0.30	0.25	0.15	-0.29	-0.20	-0.04	-0.05
Ysc				1	0.32	0.10	-0.32	-0.10	-0.25	-0.25
yx					1	0.41	-0.85	-0.19	-0.04	-0.12
yo						1	-0.37	-0.67	-0.10	-0.18
ms							1	0.54	0.03	0.07
mo								1	0.03	0.06
C1									1	0.98
C										1

Figure 5.5: Correlation coefficients for the ten parameter distributions after 100 bootstrap samples. The greatest correlation coefficients are highlighted in bold.

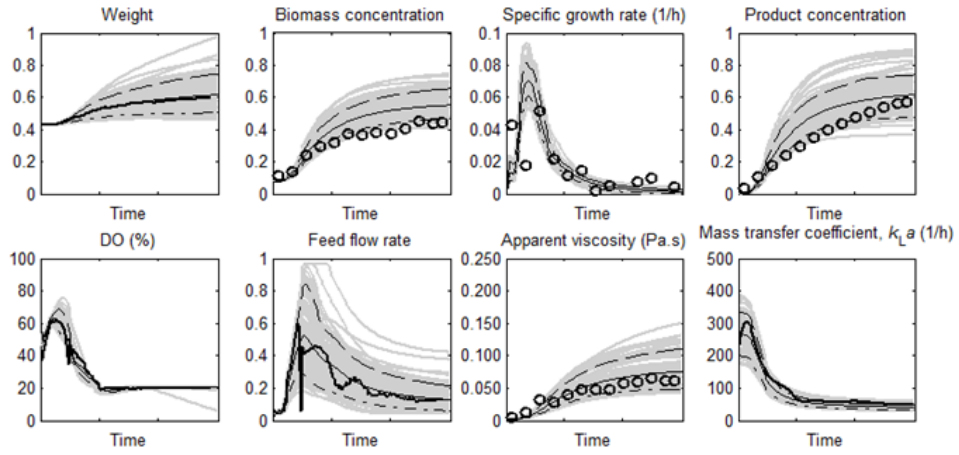
Ysp but inversely proportional to Ysx. This means that an increase in one is simply balanced by an increase in the other.

The stoichiometric yield parameters, yxo and yxs, in the model have a low level of uncertainty, despite them being assigned the greatest uncertainty level in the expert review process. This was since the parameters are difficult to determine, since the experiments are carried out at low growth rates. It has been shown by the bootstrap method however, that the average parameter values are well defined for this process. They also have no highly significant correlation to other parameters and so it is concluded that these terms are well parametrised in the model.

The greatest relative error is seen in the substrate maintenance parameter, ms. This parameter was in the highest uncertainty class in the original expert review process, meaning an expected uncertainty range of 50%, however the results of the bootstrap method show not only is this parameter very uncertain, it also does not follow a Gaussian distribution. This suggests that the parameter may be poorly defined in the model, and possibly a constant for this parameter is not appropriate. The parameter, ms, is only present in calculation of the growth rate. It is unclear why the parameter distribution is so skewed.

Other than the biological parameters, there is also a strong positive correlation between C,

Expert review parameter uncertainty



Bootstrap determined parameter uncertainty

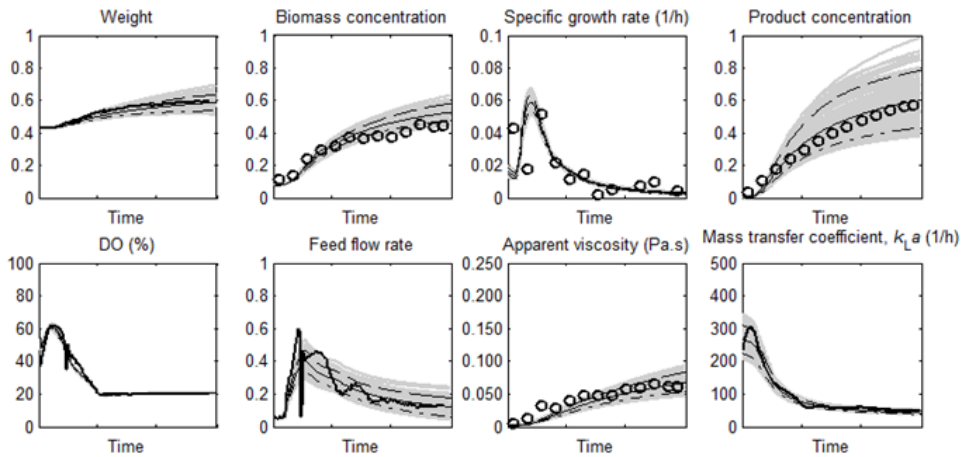


Figure 5.6: 100 Monte Carlo simulations using parameter uncertainty estimate by expert review process from Albaek et al. 2011 (top) compared to parameter uncertainty obtained from bootstrap method and Latin hypercube sampling (bottom).

in the mass transfer coefficient calculation, and C1 from the viscosity prediction. From the sensitivity analysis provided in Albaek et al. (2012) it is seen that both C and C1 have positive regression coefficients for the viscosity meaning greater C and C1 values mean greater viscosity. However for the k_La calculation C has a positive correlation and C1 a negative correlation. This then shows the interplay between viscosity and k_La ; a greater viscosity means a lower k_La .

Parameter			Original	Bootstrap	Change
Y _{sx}	g DCW / g substrate	(observed)	10%	25%	+15%
Y _{sp}	g product / g substrate	(observed)	10%	24%	+14%
Y _{so}	g oxygen / g substrate	(observed)	10%	77%	+67%
Y _{sc}	g CO ₂ / g substrate	(observed)	10%	57%	+47%
y _{xs}	g substrate / g DCW	(stoichiometric)	50%	11%	-39%
y _{xo}	mol O ₂ / g DCW	(stoichiometric)	50%	12%	-38%
m _s	g substrate / g DCW / h	(maintenance)	50%	114%	+64%
m _o	g O ₂ / g DCW / h	(maintenance)	50%	34%	-16%
C1	constant for viscosity		30%	28%	-2%
C	constant for k _L a		30%	15%	-15%

Figure 5.7: Summary of parameter uncertainty estimates.

5.4 Conclusions and future work

Mechanistic models are applicable to many stages of fermentation process development, and also provide a summary of the process understanding. For this reason they are valuable tools for pilot scale fermentation studies. In this chapter, an established mechanistic process model is analysed using uncertainty analysis methods in order to gain an understanding of the parameter uncertainty estimates.

The choice of parameter estimation method is shown to have a great effect on the resulting model uncertainty, which gives a different conclusion about the model robustness. It is not the purpose to conclude on the most suitable approach to parameter uncertainty estimates, but to compare different methods for the purpose of understanding the model. The bootstrap method gave a lower model uncertainty than was originally described in Albaek et al. (2011), despite the fact that some parameter uncertainty estimates were increased with the bootstrap method, whilst others were decreased. This is also due to the consideration for the parameter correlations, which reduces the parameter space which is sampled. The model parameter set provided by the bootstrap method provided an improved model prediction.

Figure 5.6 compares the expert review parameter uncertainty and the bootstrap defined parameter uncertainty by simulating both using Monte Carlo simulations after Latin hypercube sampling. The uncertainty estimate obtained from the bootstrap method is lower in all model outputs, and especially for the weight, biomass, viscosity, and feed flow rate predictions. The method of parameter fitting has also resulted in an improved parameter set, as shown by comparing the average model prediction in the top of Figure 5.6 with the lower figure. Especially with the product prediction, the average model prediction is much improved. This indicates that a process of mechanistic modelling followed by parameter fitting by bootstrap sampling is suitable for obtaining a set of model parameters, and analysing the resulting parameter distributions. It is also highly valuable for future model development, that the parameter distribution and correlations are understood. This could lead to better parameter estimates, and lead to identification of areas for model improvement.

The uncertainty analysis shows that the greatest model uncertainty is in the product concentration prediction. This is due to the high uncertainty in the maintenance terms, and the observed yield coefficients for both product and biomass, since the product formation is also

dependent on the growth rate of biomass as shown in equation 5.11. In this way the uncertainty in the biomass concentration has an effect on the uncertainty in the product concentration. It may be that the assumption of growth associated product formation is not valid, and this area of the model should be a focus of future work. It is shown that the maintenance term for substrate consumption is poorly defined in the model, and it would be of interest to focus on the use of the yield parameter and maintenance terms, to find a structure where the parameter estimates showed a lower uncertainty and a normal distribution.

The substrate balance should also be considered in future application of the model. Currently, there is assumed to be no accumulation of substrate, and the feed rate is solved by application of a PI controller. The data shows the fluctuating feed rate as it is applied for control of the dissolved oxygen concentration. In the model all feed is consumed at every time instance, and therefore it is not possible to recreate the reality which is shown, where by the feed rate oscillates due to slight over, and under feeding. For the purpose of modelling the general trends in the data, the assumption of no feed accumulation is valid, since all the model data has been operated in well controlled conditions. However if the model is to be applied for control applications, it will be very important to account for the situation of overfeeding. If this is not considered the dynamics of the process due to an excess of substrate will not be accounted for.

Part III

Monitoring

6 Development of a stoichiometric state estimator

Chapter 5 describes a robust mechanistic model which has previously been developed to describe the pilot scale fermentation process operated at Novozymes A/S [Albaek et al., 2011, Albaek et al., 2012]. In this chapter, the model is applied on-line, and therefore there is available measurement data which may be utilised in order to improve the model prediction. This chapter of the thesis is based upon the following article:

Application of a Mechanistic Model as a Tool for On-line Monitoring of Pilot Scale Filamentous Fungal Fermentation Processes- The Importance of Evaporation Effects

Mears L., Stocks S. M., Albaek M. O., Sin G., Gernaey K. V.

Biotechnology and Bioengineering 114, 3: 589-599. (2017)

6.1 Introduction

Chapter 5 describes a robust mechanistic model [Albaek et al., 2011, Albaek et al., 2012] which has previously been developed to describe the pilot scale fermentation process operated at Novozymes A/S. The model includes both a physical process description with mass transfer relations, as well as a description of the metabolic rates in the system by means of yield parameters and maintenance terms for substrate consumption. Although the model describes the trends seen in the data, it has been shown in section 5.3 that there is a high level of uncertainty in some of the parameters used in the mechanistic model, and this uncertainty mainly concerns maintenance and yield parameters. The term describing the substrate consumed for maintenance is not well defined and has an unacceptably high uncertainty. The greatest relative error of prediction is in the product concentration, as shown in Figure 5.6, which is related to the uncertainty in the yield parameters. It is considered that the model describes the physical processes well, however the metabolic changes over the course of the fermentation are not well captured. This is due to a lack of understanding of how the physical environment affects the metabolic activity, and this is outside the scope of the model. The prediction accuracy is very good for application to planning operations, and studying the physical system, which was the application of interest for the model when it was developed. However, it is not considered accurate enough for the purpose of on-line monitoring, which is of interest in this work.

For monitoring and control applications the model will be applied on-line. This means that there is then the opportunity to utilise available on-line measurement data, in order to improve the model prediction, by using a state estimation approach. In this case the aim of the state estimator is to predict the biomass and product concentration, as these are currently poorly defined in the model and are otherwise available only off-line. The aim is to incorporate available measurement data on-line into a state estimator, which is coupled to the physical process model. In this way it is possible to obtain a greater prediction accuracy when the model is applied on-line, which will aid process monitoring, and ultimately process control.

6.1.1 Bioprocess monitoring

Monitoring of bioprocesses requires reliable on-line measurements, however there is a lack of on-line sensors for key parameters of interest in the field, such as substrate, product and biomass concentration [de Assis and Filho, 2000, Sagmeister et al., 2013, Alves-Rausch et al., 2014], which are applicable to pilot and production scale. There are challenges specific to the development of in-line sensors for industrial fermentation systems. These include the need for the probe to be robust to sterilisation, and to be stable over long operation times [Alford, 2006]. In addition, there is also the issue of regulation, and the need to obtain approval for changes made to the hardware used in a process operating under good manufacturing practices (GMP) [Gernaey, 2015]. A practical issue is also the limited number of ports for in-line probes on the stainless steel vessels. This lack of on-line state measurement, limits the ability to monitor the progression of fermentation systems.

There are many causes of batch-to-batch variation in biological systems, based on physiological differences, metabolic shifts, or disturbances, for example small differences in raw materials [Villadsen et al., 2011]. If on-line estimates of key performance indicators, such as product concentration, were available, it allows for better process monitoring, and also allows for implementation of advanced control. With the process analytical technology (PAT) guidelines being introduced (FDA, 2004), there is also an additional drive towards increasing process knowledge and monitoring capabilities [Gernaey et al., 2010].

6.1.2 Soft sensors applied to fermentation systems

Due to the current limitations in process monitoring, there is an interest in soft sensors, which utilise on-line measured variables to predict the unknown states in real-time [Luttmann et al., 2012, Sagmeister et al., 2013]. However, a report by Luttmann et al. (2012) states that there is limited application of soft sensors in industry, despite the advantage of real-time process understanding, and the fact that there is no need for investment in additional hardware. Some of the challenges associated with implementation of a soft sensor in an industrial context are cited in the report. These include data availability for model development at industrial scales, and data quality, which may contain outliers or have issues with sensor drift. Another challenge is the requirement for additional computer hardware where there is connectivity to the on-line measured data. Finally, the need for re-calibration of the model is perceived as limiting the practical applicability. These factors must be considered in the development of a soft sensor if it is to be applied in practice.

State estimator models may be developed based on first principles understanding, data-driven methods or by hybrid modelling. Data driven methods do not require an understanding of the system, and may therefore be considered a faster approach to model development. Data-driven methods may, for example, be based upon artificial neural networks [Chen et al., 2004b, Linko et al., 1999], fuzzy logic [Araújo-Bravo et al., 2004, Luttmann et al., 2012], or multivariate statistical modelling approaches such as principal component analysis [Yuan et al., 2014, Zhang and Lennox, 2004]. These methods have the disadvantage that they are unreliable when extrapolating outside the range of the data used to develop the model. For production scales this may be less of a problem, where process conditions are generally defined. For application at pilot scale however, this is undesirable since new processing conditions are investigated. The development of a data-driven model also provides little insight into the process, since the model parameters have no physical meaning. Hybrid modelling approaches combine some of the benefits of data driven models, with the more robust basis of a mechanistic model. It is not necessary to have a

full understanding of the process of interest and therefore the model development is faster, and the resulting model may have enhanced extrapolation capabilities [von Stosch et al., 2014].

In contrast, first principles soft sensor models are based on a fundamental understanding of the system [Sagmeister et al., 2013, Sundström and Enfors, 2008]. Their development is based on existing process understanding, and the model parameters have a physical meaning. Parameter values therefore provide information about the process of interest, for example yield coefficients, which may be used to compare between strains. Soft sensor models incorporating stoichiometric balances are valuable, since they are scale independent, and based on the fundamental biochemical reactions. Since it is observed that yield coefficients change over the course of a fermentation [Golabgir et al., 2015, Jenzsch et al., 2006c], this method is interesting as it allows for this adaption within a mechanistic model structure. In addition, stoichiometric balances utilise flow rates as input variables, which then avoids the need for in-line probe measurements, which may be highly dependent on their position in the vessel, and the calibration accuracy. The mechanistic model approach also has the benefit of being more generic to new processes which may require a small adaption to the model, but in general should be applicable to different strains and processes.

This work discusses the application of a first principles soft sensor model to pilot scale filamentous fungal fermentation systems operated at Novozymes A/S. The model comprises of an on-line parameter estimation block, coupled to a dynamic model of the system. The parameter estimation block is based on a stoichiometric balance, where the current rates of product and biomass and water formation and feed consumption are identified from available off gas measurements and ammonia addition. This parameter estimate is then used as an input to a mechanistic process model, which describes the mass transfer capabilities of the system based on the operating conditions, including stirrer speed, aeration rate, headspace pressure and temperature. The model is developed and calibrated using a historical dataset as described by Albaek et al (2011). The model is then implemented at the fermentation pilot plant of Novozymes A/S and validated on-line using fourteen new batches at 550L scale, utilising a different host strain and product. With implementation of a robust soft sensor, it is possible to incorporate the state estimate into a control structure, and open up possibilities for achieving more advanced process control.

6.2 Methods

6.2.1 Model description

Adaptations to the mechanistic model

The model is an adaptation of the model developed by Albaek et al. (2012), as described in detail in Chapter 5. The fundamental changes to the model are described.

Mass basis

A fundamental change is made to the model in order to define the system on a mass basis rather than a volume basis. This factor is considered important to the model accuracy at this scale. Modern processes in industrial biotechnology cannot be treated as dilute systems. Fungal fed-batch systems are reported to produce up to 30g/L biomass [Riley et al., 2000, Tolan and Foody, 1999], and may also produce 100g/L product, in the example of *Trichoderma reesei* producing cellulases [Cherry and Fidantsef, 2003, Schuster and Schmoll, 2010]. In addition, gas hold up changes with time in a process [Hofmeester, 1988]. These factors make estimates of density and volume somewhat difficult; with densities in the range 1.05 to 1.3 kg/L there is the potential for errors in concentration in the

range 5 to 30%. This means that estimates of state based on volume measures can only be validated by accurate measurements of broth density or gas hold up, the latter being far from trivial. Based on these factors it is considered that modelling of concentrations on a unit mass basis instead of volume is a more accurate approach, as it is independent of density or gas hold up.

Evaporation rate calculation

The mass prediction accounts for the feed added, as well as the evaporation rates, which are significant in a pilot scale system of over $0.5m^3$. The rate of evaporation, F_{evap} , was a constant in the mechanistic model. This is now represented by a model equation, where the outside air temperature and relative humidity must be defined, in addition to accounting for the operating conditions. For a given batch, the evaporation rate is dependent on the air humidity and air temperature, as well as the processing conditions for aeration rate, headspace pressure and process temperature.

Yield parameters

It was shown in Section 5.3 that there was high uncertainty in some of the mechanistic model parameters. This was mainly surrounding the use of yield coefficients and maintenance terms. In order to avoid the use of these terms, a mass balance approach will be assessed, where by the rates of formation of biomass and product are determined from measured data. This method avoids the need for fixed model yield parameters or growth rates, which are typically incorporated in unstructured fermentation models. Similarly, the substrate concentration is calculated based on the feed added, and the estimated consumption rate from the state estimator. This requires a known substrate concentration in the feed.

Model structure

The model comprises of two compartments. The first unit is a stoichiometric balance which is used for on-line parameter estimation based on the available measured data. The second compartment is a dynamic fermentation model, which utilises the parameters estimated in the previous block. The output is a state vector, \mathbf{x} , and outputs, \mathbf{y} .

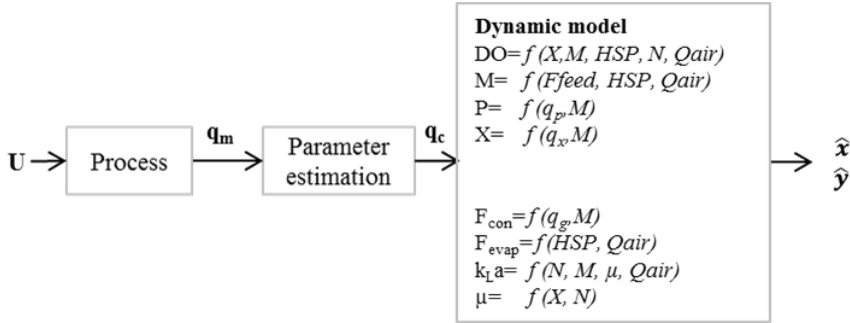


Figure 6.1: Representation of the model structure showing the parameter estimation block coupled to the dynamic process model

$$\mathbf{U} = \begin{pmatrix} CER \\ OUR \\ AMM \end{pmatrix} q_m = \begin{pmatrix} -q_n \\ -q_o \\ q_c \end{pmatrix} \mathbf{x} = \begin{pmatrix} DO \\ M \\ P \\ X \end{pmatrix} \mathbf{y} = \begin{pmatrix} k_{la} \\ \mu_{app} \\ F_{con} \\ F_{evap} \end{pmatrix}$$

One of the benefits of this model structure, is that the number of model parameters has been reduced. From the original model parameters in the full dynamic model, only two are still applied, for the $k_L a$ and viscosity prediction. This is since the other eight parameters which were yield and maintenance parameters are now replaced by on-line calculated rates of formation for biomass and product. There are additional model parameters however, which correspond to the compositions of biomass and product. The total number of model parameters is therefore eight.

Parameter estimation

The stoichiometric balance was defined using a single carbon source and assuming no additional by-products are formed.

$$q_g CH_2O + q_o O_2 + q_n NH_3 = q_x CH_{XH} O_{XO} N_{XN} + q_p CH_{PH} O_{PO} N_{PN} + q_c CO_2 + q_w H_2O \quad (6.1)$$

In Equation 6.1, q_g , q_o , q_n , q_x , q_p , q_c , and q_w are seven molar rates of formation (mol/h), and XH , XO , XN and PH , PO , and PN are the hydrogen, oxygen, and nitrogen molar quantities of the biomass and product respectively. This equation can be solved by use of a carbon, nitrogen, hydrogen and oxygen balance.

Given that there are four equations and seven parameters, we must define three terms. This will be the q_o , q_n , and q_c . In a pilot scale operation, q_n is a measured flow rate and q_o and q_c are calculated in real time from high quality gas measurements to give the oxygen uptake rate, OUR and the carbon dioxide evolution rate, CER . It is not desirable to use the substrate feed rate as an input as it is subject to batch-to-batch variations in concentration.

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 2 & 0 & 3 & XH & PH & 0 & 2 \\ 1 & 2 & 0 & XO & PO & 2 & 1 \\ 0 & 0 & 1 & XN & PN & 0 & 0 \end{pmatrix}$$

$$\mathbf{q} = \begin{pmatrix} -q_g \\ -q_o \\ -q_n \\ q_x \\ q_p \\ q_c \\ q_w \end{pmatrix}$$

Given \mathbf{E} and a column vector of rate coefficients, \mathbf{q} , the solution is found as $\mathbf{E}\mathbf{q}=\mathbf{0}$. Since we have defined three measured rates, we can instead define the problem by separating the measured rates, \mathbf{q}_m , and the calculated rates, \mathbf{q}_c .

$$\mathbf{E}_m \mathbf{q}_m + \mathbf{E}_c \mathbf{q}_c = 0 \quad (6.2)$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 3 & 0 & 0 \\ 0 & 2 & 2 \\ 1 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} -q_n \\ -q_o \\ q_c \end{pmatrix} + \begin{pmatrix} 1 & 1 & 1 & 0 \\ 2 & XH & PH & 2 \\ 1 & XO & PO & 1 \\ 0 & XN & PN & 0 \end{pmatrix} \times \begin{pmatrix} -q_g \\ q_x \\ q_p \\ q_h \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} -q_g \\ q_x \\ q_p \\ q_h \end{pmatrix} = -(\mathbf{E}_c)^{-1} \mathbf{E}_m \mathbf{q}_m$$

Dynamic model compartment

Rates of formation The molecular mass of all species in the system can be defined. This means we have a fully characterised stoichiometric balance of the system, which can now be used as a state estimator, and to solve for other physical parameters. This is based on a known composition of biomass and product. These parameters are referred to as XH , XO , XN , PH , PO and PN .

$$\begin{pmatrix} Mr_g \\ Mr_o \\ Mr_n \\ Mr_x \\ Mr_p \\ Mr_c \\ Mr_w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 2 & 0 & 3 & XH & PH & 0 & 2 \\ 1 & 2 & 0 & XO & PO & 2 & 1 \\ 0 & 0 & 1 & XN & PN & 0 & 0 \end{pmatrix}' \times \begin{pmatrix} 12 \\ 1 \\ 16 \\ 14 \end{pmatrix}$$

Mass balance The current state of biomass and product is calculated in terms of concentration with respect to mass rather than volume. The differential change in mass is calculated as a mass balance formulation.

$$\frac{dM}{dt} = \frac{S_F F}{1000} - F_{evap} + \frac{q_o Mr_o}{1000} - \frac{q_c Mr_c}{1000} \quad [kg/h] \quad (6.3)$$

The rate of evaporation is not defined as a model constant, but instead a function of the pressure and airflow rate as well as the current air relative humidity and temperature. In this work, the air conditions are defined once for a batch operation, and are assumed constant over the batch. If available, it is desirable to have an on-line input of these air measurements.

$$F_{evap} = \left[\frac{P_{out}^* Mr_h}{1000 RT} \frac{P_{atm}}{(P_{atm} + HSP)} - \frac{P_{in}^* Mr_h}{1000 RT} \right] 0.06 Q_{air} \quad [kg/h] \quad (6.4)$$

Biomass and product formation It is then possible to solve for the system states of biomass concentration, X (g/kg) and product concentration, P (g/kg).

$$\frac{dX}{dt} = \frac{q_x Mr_x}{M} - X \frac{dM}{dt} \quad [g/kg.h] \quad (6.5)$$

$$\frac{dP}{dt} = \frac{q_p Mr_p}{M} - P \frac{dM}{dt} \quad [g/kg.h] \quad (6.6)$$

Dissolved oxygen concentration

$$\frac{dDO}{dt} = k_L a \frac{(DO_{in}^* - DO) - (DO_{out}^* - DO)}{\log \frac{(DO_{in}^* - DO)}{(DO_{out}^* - DO)}} - \frac{q_o}{M} \frac{dM}{dt} \quad [mol/kg.h] \quad (6.7)$$

Substrate balance

$$\frac{dG}{dt} = \frac{S_F F - q_g Mr_g}{M} - G \frac{dM}{dt} \quad [g/kg.h] \quad (6.8)$$

6.2.2 Implementation algorithm

Figure 6.2 shows a detailed description of the on-line implementation algorithm. The on-line data is read directly from the OPC server every 30 seconds using MATLAB® timer objects. The parameters are updated every 5 minutes, using the input data which is simply averaged, assuming Gaussian noise. The updated parameters are input to the dynamic model which is then also solved using this 5 minute window. The new state estimate is then overwritten as the initial condition for the next modelling iteration. This results in a low computational demand, by only modelling at five minute intervals. The user may plot the results at any time, and the results will be updated to within a five minute sampling interval. For model calibration, a dataset was used whereby the on-line data was available at one hour intervals. This means that the same implementation algorithm was applied, except for the parameter update occurring once every hour instead of once every five minutes.

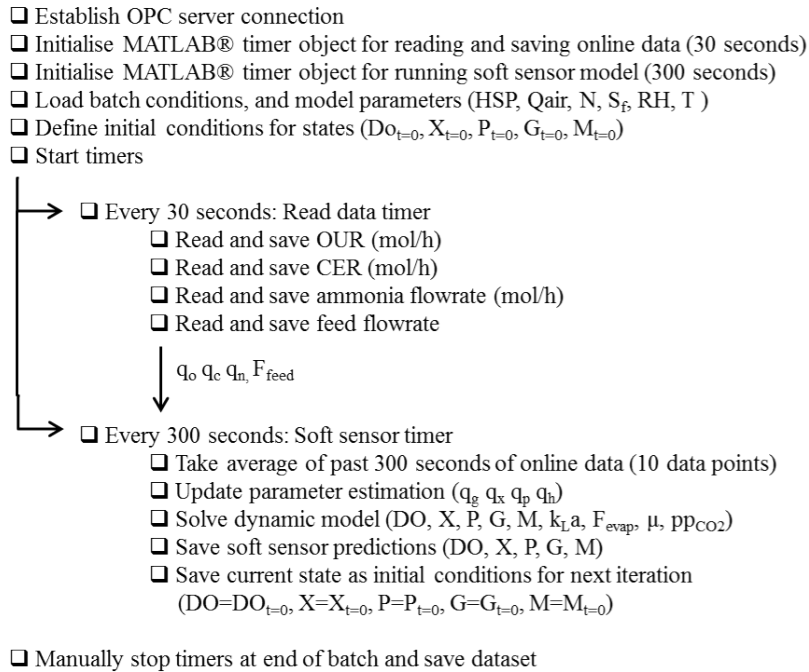


Figure 6.2: Implementation algorithm for the mechanistic model-based monitoring tool using MATLAB® timer objects

6.2.3 Off-line sample analysis

The biomass concentration was measured by dry mass determination, by drying at 105 °C for 48 hours. The biomass sample was washed twice with deionised water to remove soluble media components. The product concentration was determined based on a generic protein assay used at Novozymes A/S. Viscosity was measured off-line in an AR-G2 rheometer from TA instruments using a vane-and-cup geometry. The vane consists of four blades at right angles (14 mm x 42 mm), the cup had a 15 mm radius, and the gap between the vane and cup was 4000 μm.

Measurements were taken in the interval of 10 to 600 1/s and the Bingham plastic model was applied to describe the rheological behaviour [Bingham, 1916]. The shear rate for apparent viscosity determination was found by the approach of Metzner and Otto, ksN , where ks is 11. The k_La was determined by the direct method [Villadsen et al., 2011] also assuming a log mean driving force, as described in equation 6.7.

6.2.4 The historic dataset

The stoichiometric model parameters (XH, XO, XN, PH, PO, PN) were obtained by least square fitting to a historical dataset of eleven batches. This dataset is designed as a full factorial design including two levels for three process variables, namely specific power input (1.5-15kW/m³), aeration rate (96-320NL/min) and headspace pressure (0.1-1.3 bar) as described by Albaek et al. (2011). The processing conditions affect the biomass concentration and product concentration achievable, due to oxygen mass transfer limitations. This makes this dataset ideal for calibration of the model parameters, as there is significant deviation in these states between the batches.

6.2.5 Validation batches

The model was then validated on-line for fourteen new batches. These batches are also operated at different stirrer speeds, aeration rate and headspace pressure which is the reason for the different biomass and product concentrations achieved. The conditions were mostly within the design space as described by Albaek et al. (2011), however some are also just outside of this range. In this newer process, media optimisation has led to the inclusion of partially soluble compounds in the media. Due to the solid content in the media, it has not been possible to quantify biomass concentration in these batches.

6.2.6 Statistical assessment of model fit

When assessing the goodness of fit for the model, the root mean sum of squared errors (RMSSE) was applied, as defined in equation 6.2.6, where $y_{(meas,i)}$ is one of, n , measurement points, and \hat{y}_i is a model prediction of the same variable. This is expressed as a percentage of the average measured value, for confidentiality reasons. In addition, to assess the calibration model fit compared to the validation model fit, the Janus coefficient, J^2 , was used, as discussed in Sin et al. (2008). In this work, the model prediction of product concentration is the primary focus, as this is a key process performance indicator which otherwise takes considerable time to obtain analytically.

$$RMSSE = \frac{\sqrt{\frac{1}{n} \sum_i^n (y_{meas,i} - \hat{y}_i)^2}}{\frac{1}{n} \sum_i^n y_{meas,i}} \quad [\%] \quad (6.9)$$

$$J^2 = \frac{\frac{1}{n_{val}} \sum_i^{n_{val}} (y_{meas,i} - \hat{y}_i)^2}{\frac{1}{n_{cal}} \sum_i^{n_{cal}} (y_{meas,i} - \hat{y}_i)^2} \quad [-] \quad (6.10)$$

6.2.7 Parameter estimation uncertainty

In order to conclude on the uncertainty of the fitted stoichiometric parameter values, bootstrap sampling is applied [Efron, 1979], as described in more detail in Section 5.3.2. In this method, the residuals between the model and the data are sampled in order to create simulated datasets for fitting. The errors are sampled from the biomass concentration, the product concentration, the dissolved oxygen concentration and the mass. The method for sampling is random sampling

of 100 residuals with replacement. By fitting to each of the simulated datasets, a distribution of parameter values is identified, which provides an indication of the parameter uncertainty. The parameter uncertainty in this work is provided as two standard deviations as a percentage of the mean.

6.3 Results and discussion

6.3.1 Calibration using historic dataset

The soft sensor model was applied off-line to an eleven batch pilot scale dataset, as described in Section 6.2.4. The purpose was to fit the model parameters for the stoichiometric model (XH, XO, XN, PH, PO, PN). The measured on-line data for carbon evolution rate (q_c), oxygen uptake rate (q_o) and ammonia addition rate (q_n) are used as input to the parameter estimation block in order to simulate the system as would be done on-line. The parameter update occurs every hour.

Figure 6.3 shows the results of the dynamic model for one batch of data. This batch was chosen as it shows measured data which has previously been published (Albaek et al., 2011). There is a very good agreement between the model prediction and the measured data for all variables. The dynamics in the dissolved oxygen profile as shown in Figure 6.3 are due to the oxygen uptake rate. This measured oxygen uptake rate has fluctuations corresponding to the feed rate applied, however in the real system this direct impact is not seen. This suggests a limitation in the model description which should be considered when the model is applied. In addition, the final substrate concentration is seen to go negative, which may be explained by uncertainty in the initial substrate concentration in the batch phase, variations in the feed concentration between batches, as well as consumption of additional media components which are not considered in the stoichiometric balance methods. This is a limitation in the model, as of course this is not physically possible. The consumption rate of substrate is independent of the substrate concentration, as it is entirely based on the stoichiometric balance. The substrate balance is certainly an area of future work for the model. Overall, the qualitative trends are captured well in the model, and the prediction accuracy is considered acceptable.

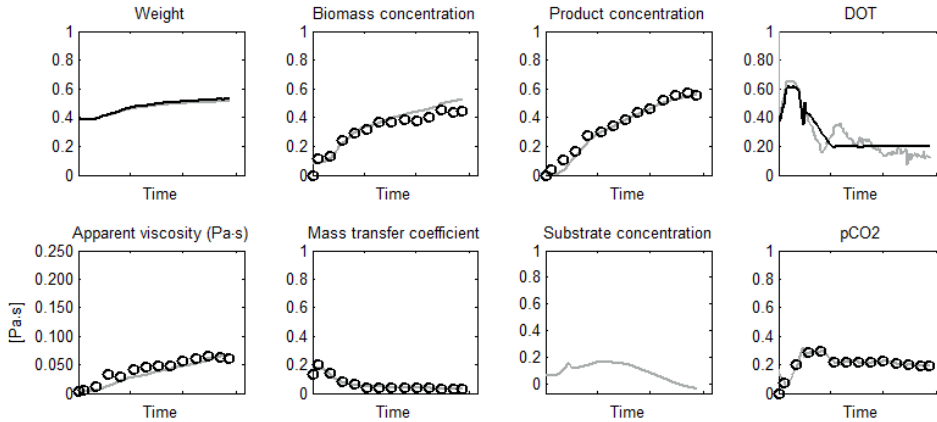


Figure 6.3: Model prediction for one calibration batch. Data scaled for confidentiality reasons.

The product concentration prediction and the biomass concentration prediction for all eleven batches is shown in Figure 6.4. Across the eleven batches, the mean deviation between the

model prediction and the data for product concentration ranges from 4.3% to 26.2%, and a summary of the results for all 11 batches is shown in Table 6.1. The mean deviation between the model prediction and the data for biomass concentration ranges from 6.2% to 21.8%. The model results are considered robust to the different operating conditions, and the accuracy of the model prediction across the batches operated in such different physical conditions is very encouraging for future application as a monitoring tool.

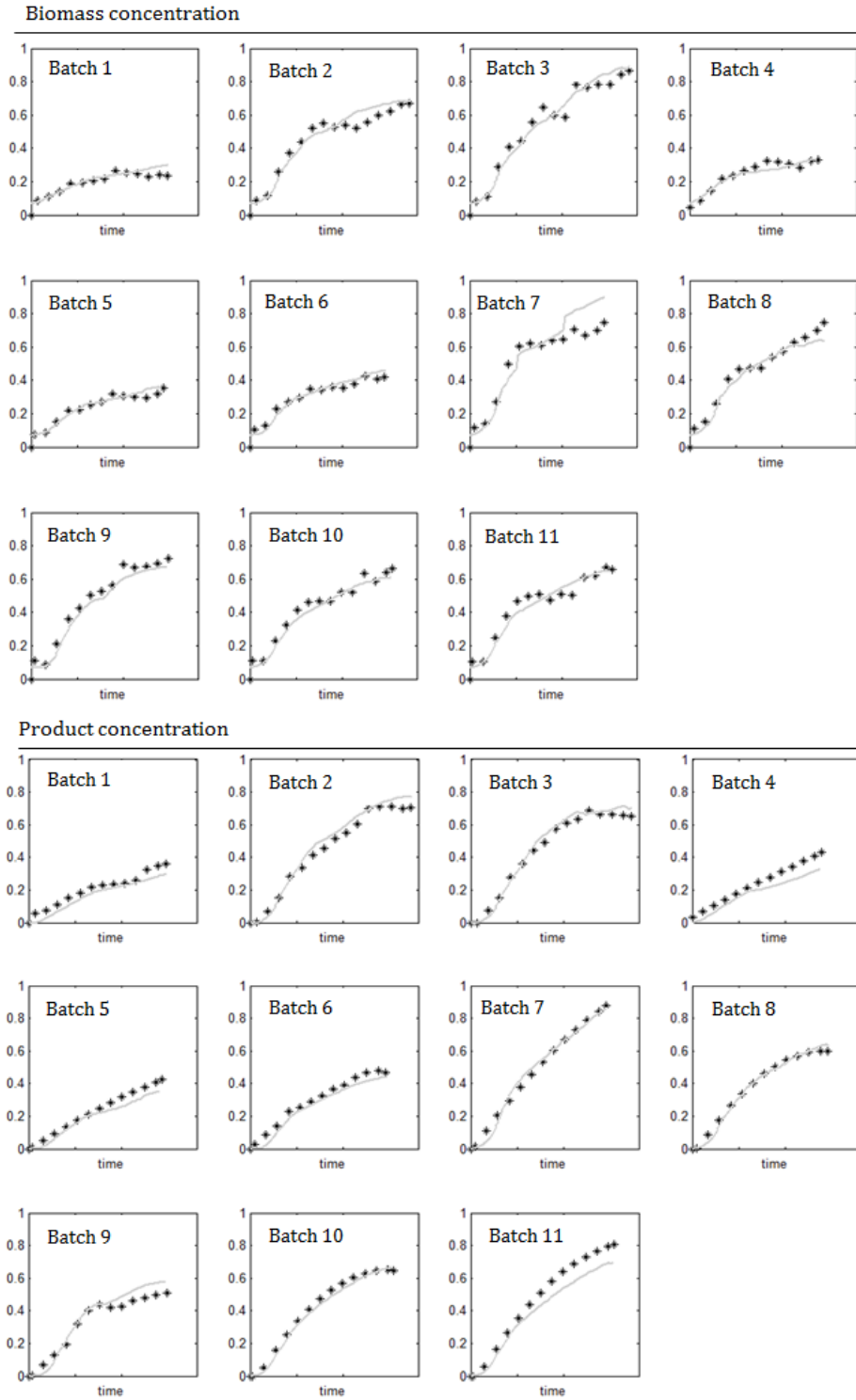


Figure 6.4: Biomass concentration prediction (top) and product concentration prediction (bottom) for 11 batches of pilot scale data, as discussed in Albaek et al. 2011

6.3.2 On-line validation in Novozymes A/S pilot plant

Once the model parameters have been fitted to the calibration batches, it is possible to implement the model on-line. Focusing on the product concentration as the fundamental measure of process performance, it is shown that the model predicts the current product concentration well, with an average RMSSE of 16.6% in fourteen new validation batches. Table 6.1 summarises the assessment of model fit for the validation batches. It is seen from Table 6.1 that the Janus coefficient is close to one for all batches, which suggests that the model accuracy in the validation batches is comparable to that of the calibration batches, with an average Janus coefficient of 1.5. Batches 8 and 11 show the greatest Janus coefficient, due to a significant underprediction of the product concentration at the end of the batch. The high Janus coefficient in Batch 6 may be partly due to measurement error, as it is seen one data point does not follow the trend. The greatest RMSSE is in Batch 2. Although the absolute error of the prediction is comparable to the other batch predictions as shown by the Janus coefficient, the percentage error is greater due to the lower average product concentration. This shows the importance in utilising more than one assessment of model fit, where in this case the Janus coefficient shows that the model fit is no worse based on absolute errors.

Table 6.1: Statistical analysis of model fit of product concentration for 11 calibration batches, and 14 on-line validation batches. Root mean sum of squared errors, RMSSE (%), and the Janus coefficient, J^2 is shown.

Calibration	1	2	3	4	5	6	7	8	9	10	11					mean
RMSSE (%)	21.8	6.4	6.0	26.2	10.7	8.7	10.3	11.8	24.4	4.3	14.5					13.2
Validation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	mean	
RMSSE (%)	20.5	48.4	8.8	9.3	14.5	18.3	11.9	16.6	12.1	16.7	23.3	15.5	5.4	11.4	16.6	
J^2	1.3	1.1	0.4	0.5	2.0	2.7	0.7	3.5	1.7	1.5	3.5	1.5	0.1	0.6	1.5	

Overall the results show that the model provides acceptable prediction accuracy of product concentration, and that the model is calibrated successfully, such that the model errors seen in the validation batches are of a similar degree to those obtained from the calibration set. An on-line measure of product concentration is a valuable monitoring parameter which allows operators to compare between batches on-line, and intervene in the case of poor performance. The complete model output is shown in Figure 6.5 for one of the fourteen additional batches, which was selected since the results are comparable to the results reported in the calibration batch. This full model output provides important monitoring information to operators, including the viscosity measurements, which are of high importance in industrial filamentous fungal processes [Olsvik and Kristiansen, 1994].

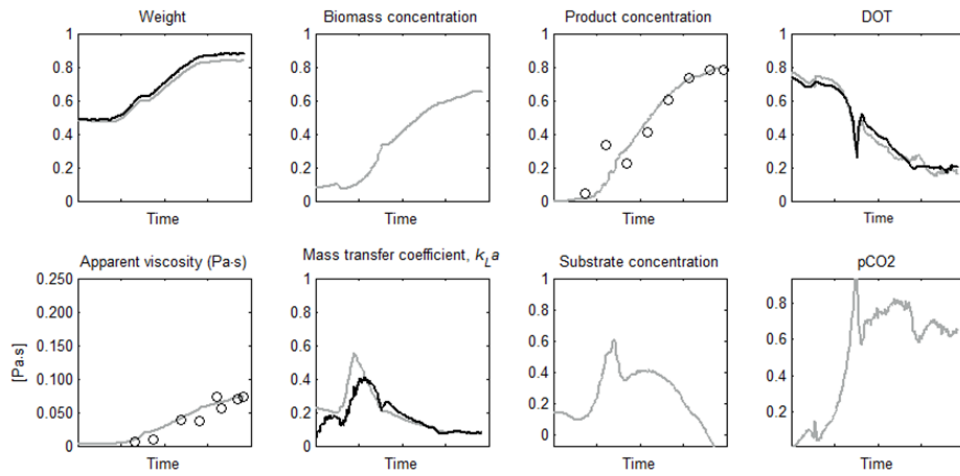


Figure 6.5: Coupled parameter estimator and dynamic model applied on-line in Novozymes A/S fermentation pilot plant. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons.

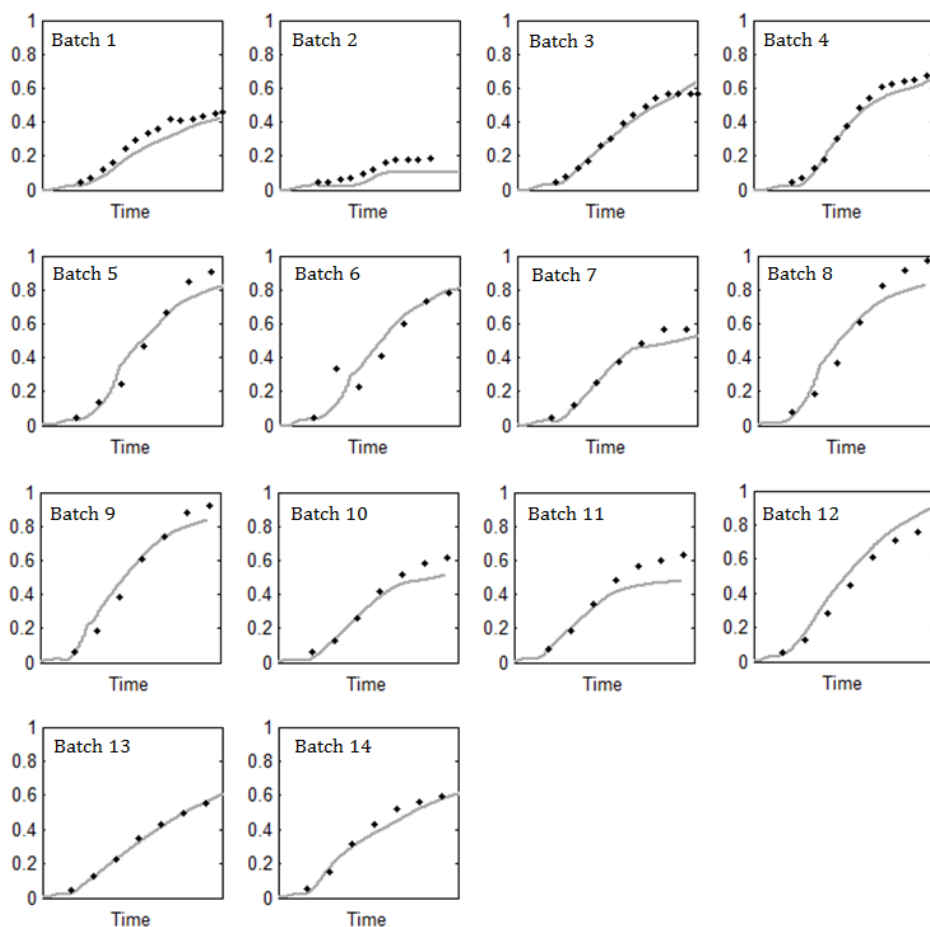


Figure 6.6: Product concentration prediction for fourteen validation pilot scale batches. Batches were performed using different operating conditions for headspace pressure, aeration rate and stirrer speed, resulting different levels of product formation. Model prediction (grey), off-line measured data (black). Scaled axis labels for confidentiality reasons.

Evaporation

It is also important to note that the mass prediction model is accurate, utilising the evaporation term in the model. This is an important addition to a pilot scale fermentation model, which is not often discussed in literature focussed on smaller scale experimentation. The evaporation rates calculated are significant, when considering fungal fermentation processes are operated for approximately one week. At the maximum rate of 40 kg/week, this equates to roughly 10% of the final mass at this scale over a week operation. The complication of implementing an evaporation model is that it depends on the relative humidity and temperature of the incoming air, as this defines the water content incoming to the system, in addition to the air flow rate, headspace pressure and temperature of the system, which defines the rate of water stripping. The environmental conditions are seen to vary significantly over a year period, as shown in Figure 6.7, and this fact must be accounted for in the model. Currently this is done by manually changing air inlet conditions, but this could become integrated as a measured parameter also, which is part of the future work for this model.

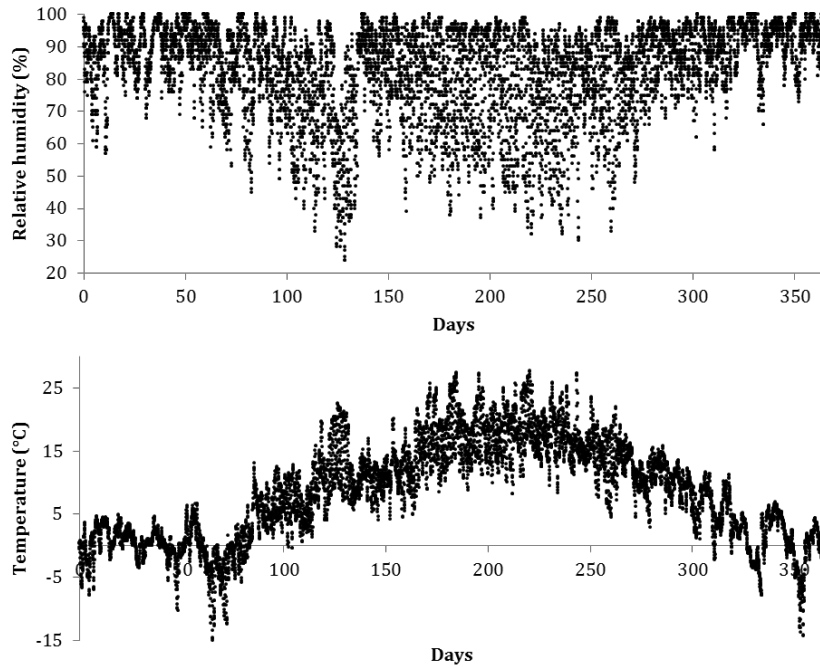


Figure 6.7: Temperature and relative humidity data obtained from the 2001-2010 Design Reference Year for Denmark dataset from the Danish Meteorological Institute [Grunnet Wang et al., 2013]. The correlation coefficient is -0.53.

It is also seen in Figure 6.7 that the temperature of the air and the relative humidity of the air are correlated, with a correlation coefficient of -0.53. With this understanding of the expected variation in inlet conditions, it is possible to simulate the range of expected evaporation rates over a year period. Latin Hypercube Sampling [Helton and Davis, 2003] is used to simulate 250 data points which represent this year of data, taking into consideration the correlation in input variables using the method of Iman and Conover [Iman and Conover, 1982, Sin et al., 2009].

These 250 sample points were then simulated by Monte Carlo simulations in order to show the variation in evaporation rate, for a fixed set of processing conditions, as shown in Figure 6.8. This shows the importance of accounting for evaporation in this pilot scale model, and further work will include validating the different evaporation rates over the year by applying the state estimator and assessing the mass prediction.

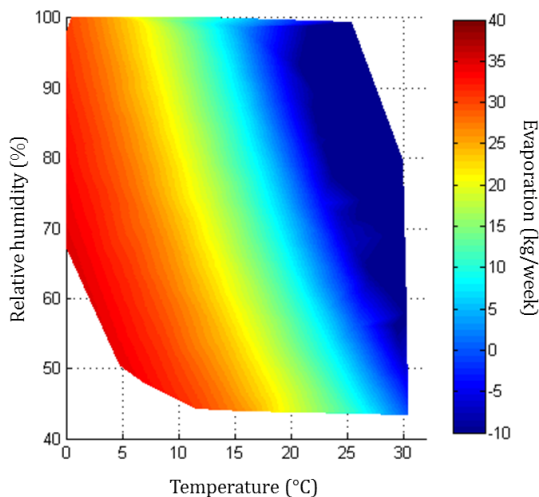


Figure 6.8: Evaporation rate model, for a pilot scale process operated at 300NL/min and 0.7 bar headspace pressure. Effect of the inlet air relative humidity and temperature is shown. The Monte Carlo simulations are applied to 250 sample points, which are generated using Latin Hypercube Sampling (Helton and Davis, 2003) taking into consideration the correlation coefficient of -0.53 which is obtained from the data in Figure 6.7. Coloured surface plot of the results is shown, with colours corresponding to evaporation rate (kg/week).

6.3.3 Model uncertainty analysis

In order to understand the uncertainty of the model prediction we assess two scenarios: The effect of measurement quality on the model prediction, and an assessment of model parameter uncertainty.

The effect of measurement quality

Since this modelling method relies on process measurements as input parameters it is important to understand how the accuracy of these measurements affects the model prediction. In Table 6.2, a constant 5% deviation is applied to the input measurements, and the effect on the model output is shown, for the same batch as shown in Figure 6.5. Table 6.2 shows the results for the relative change in the final model prediction compared to the base case for four scenarios; 5% increases are applied in each of the input parameters individually, as well as a simultaneous 5% error in both the CER and OUR. This is also assessed since these off gas measurements are utilising the same equipment, and therefore it is likely that there could be an equal error in both measurements. This may be, for example, due to an incorrect gas flow rate measurement, which affects both the OUR and CER calculation equally. The results show that the balance between the OUR and the CER is very important for the prediction, meaning that if there is an error in both CER and OUR, then the prediction is not significantly affected, however an error in only one of them means that unacceptable errors are observed. This shows the importance of the respiratory quotient ($RQ = CER/OUR$) for the model prediction, especially for the biomass prediction. The deviation in the ammonia flow rate affects both the final biomass concentration and final product concentration by less than 10%.

Table 6.2: Sensitivity analysis of the model prediction to process measurement errors. The results are given for the batch shown in Figure 6.5. The table provides the relative change in the model prediction for biomass concentration and product concentration compared to when no measurement error is applied.

	+5% NH3	+5% CER	+5% OUR	+5% CER +5% OUR
Change in final biomass prediction	3.1%	63.8%	-58.4%	5.5%
Change in final product prediction	9.3%	-10.7%	8.9%	-1.3%

Model parameter uncertainty

In addition to the model sensitivity to the input measured data, it is also important to assess the confidence in the fitted parameter values. In this case, the bootstrap method was chosen in order to provide an indication of model parameter uncertainty. The states are used for sampling model errors, meaning model outputs for weight, biomass concentration, product concentration and dissolved oxygen concentration are sampled. The results of the sampling are shown in Figure 6.9.

The parameter uncertainty can be analysed by the distribution of the parameter values, where in this case we consider two standard deviations as a percentage of the mean, for confidentiality. The bootstrap method provides a mean, standard deviation and correlation matrix for a parameter set, as shown in Table 6.3. In this case, the parameter set is the composition of product and biomass, given by the parameters PH , PO , PN , XH , XO , and XN . When all six stoichiometric parameters are fitted simultaneously, it is found that the parameter uncertainties

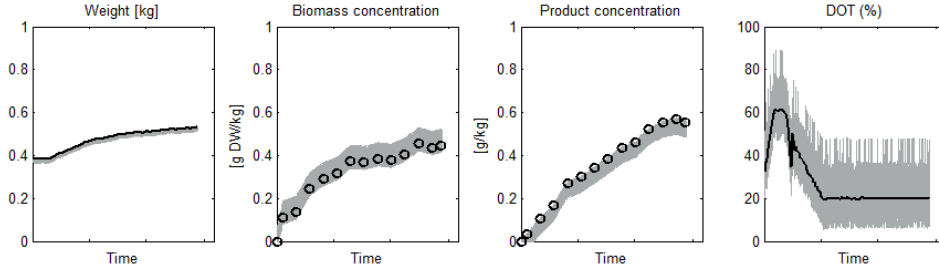


Figure 6.9: 100 bootstrap samples of model residuals for the states

for parameters XH , XO , XN , PH , PO , and PN are 16.6%, 15.6%, 41.4%, 7.2%, 16.6%, and 5.9% respectively.

There is a low uncertainty in all parameters, other than XN where 2 standard deviations is 41% of the mean value. The greatest correlation between parameters is seen with XN and PN , and XN and PO . This means that not only is there the greatest uncertainty in the XN parameter it is also highly correlated to other parameters. This is related to the model structure, since parameter estimation is based on the stoichiometric balance. There is a single nitrogen source, and so the nitrogen content of the biomass is therefore highly influential in the model, since it defines the distribution of nitrogen source between biomass and product. This is shown by the strong negative correlation between XN and PN .

The relatively high uncertainties are due to high correlation between the parameter values, which is expected in this simple stoichiometric model. If, for example, the product stoichiometry is fixed, and only the stoichiometric parameters for the biomass are fitted, the uncertainty is below 5.5% for all parameters, as shown in Table 6.4.

Table 6.3: Parameter distribution and correlations for product and biomass stoichiometry based on 100 bootstrap samples

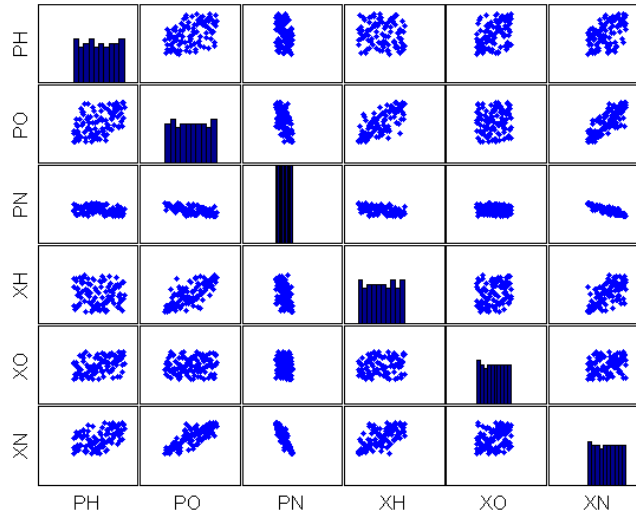
	PH	PO	PN	XH	XO	XN
Mean	0.9	0.2	0.2	1.8	0.3	0.1
Std	0.03	0.02	0.01	0.15	0.02	0.03
2 std/mean	7.2	16.6	5.9	16.6	15.6	41.4
PH	1	0.46	-0.26	-0.10	0.39	0.49
PO		1	-0.58	0.75	0.09	0.81
PN			1	-0.50	-0.19	-0.91
XH				1	0.24	0.63
XO					1	0.29
XN						1

Table 6.4: Parameter distribution and correlations for biomass stoichiometry based on 100 bootstrap samples

	XH	XO	XN
Mean	1.9	0.3	0.1
Std	0.01	0.004	0.003
2 std/mean	0.9	2.8	5.4
XH	1	0.27	-0.05
XO		1	-0.73
XN			1

Uncertainty analysis: Monte Carlo method

The uncertainty obtained from the bootstrap method can be investigated using Monte Carlo simulations in order to determine the effect of the parameter uncertainty and correlation on the model outputs. The standard deviations, and correlations from Table 6.3 (X and P parameters) and Table 6.4 (X parameters only) will be used for this purpose. Latin hypercube sampling [Helton and Davis, 2003] is applied for determining the parameter space as shown in Figure 6.10, which is simulated by Monte Carlo simulations [Metropolis and Ulam, 1949]. The methods are implemented as described by Sin et al. (2009), to give the model uncertainty as shown in Figure 6.11.

**Figure 6.10:** Latin hypercube sampling using bootstrap results

The results of the Monte Carlo simulations are shown in Figure 6.11 for when the biomass and product uncertainties are applied, and in Figure 6.12 when only the biomass composition was analysed by bootstrap sampling. Figure 6.11 shows that there is a very high uncertainty in the biomass concentration and there is a failure in the model for the dissolved oxygen prediction,

giving in some cases negative dissolved oxygen concentrations. The high uncertainty in the apparent viscosity is a propagation of the errors in the biomass prediction, as shown in Equation 5.5. The model is shown to be very sensitive to the parameter estimate. In contrast, the very low parameter uncertainty obtained by only fitting the biomass composition shows very low model uncertainty estimates.

It is valuable to conduct such an uncertainty analysis in order to understand the model structure, and the interactions of the parameters during fitting. It is now known that the nitrogen content of the biomass is the key parameter for model fitting, as it showed the greatest variability when the data was sampled. It is also known that the model fit is reliant on the balance between the biomass composition and the product composition.

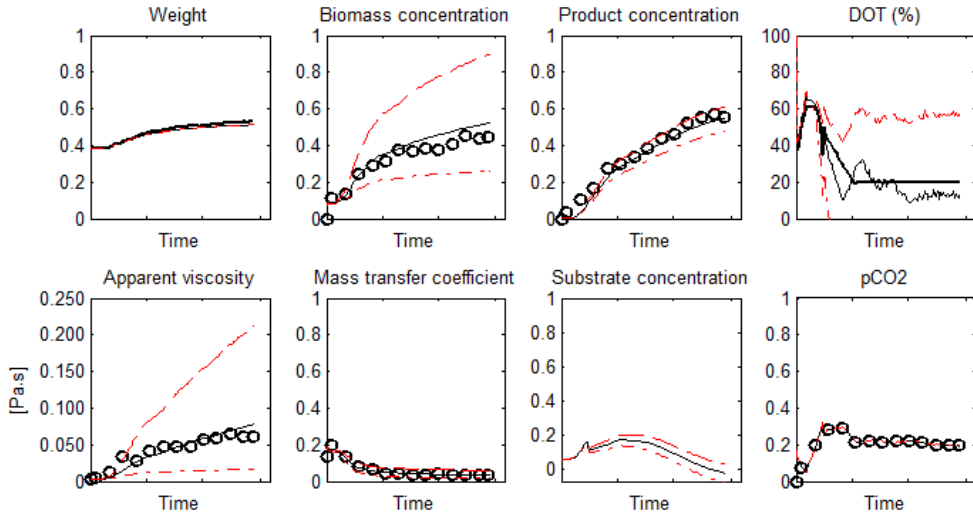


Figure 6.11: 100 Monte Carlo simulations using the the mean, 2 standard deviations and correlation matrix from the bootstrap method for PH, PO, PN, XH, XO and XN. Sampling is implemented using Latin hypercube sampling with correlation matrices. The 95 percentile is shown in red dashed lines.

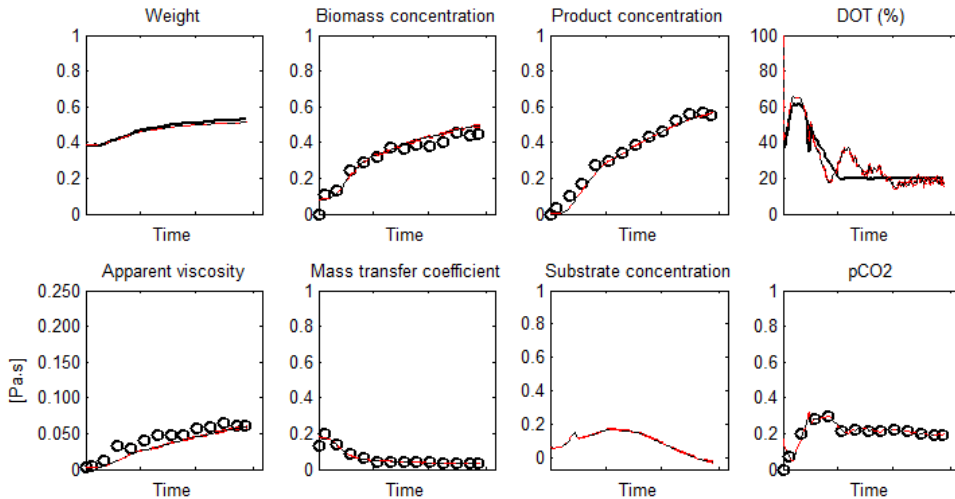


Figure 6.12: 100 Monte Carlo simulations using the the mean, 2 standard deviations and correlation matrix from the bootstrap method for XH, XO, and XN. Sampling is implemented using Latin hypercube sampling with correlation matrices. The 95 percentile is shown in red dashed lines.

6.4 Conclusions and future work

In this work, a soft sensor has been developed for 550L filamentous fungal fermentations operated at Novozymes A/S. The parameter estimation uses only standard on-line measurements, of oxygen uptake rate, carbon dioxide evolution rate and ammonia flow rate, which are considered robust measurements, not subject to drift, and without the need for calibration. It is therefore considered that the model should be applicable to other strains, other scales, and other processes, and this provides an area for future work. The flexibility of the method and the simplicity of implementation make this a valuable tool for industrial application.

It is considered important to introduce the evaporation term into the model, and define the states on a mass basis, rather than a volume basis. This accuracy in the mass prediction avoids large errors propagating to the state estimates, and aids prediction accuracy. Further future work includes validating the evaporation rate model over the year under different air conditions.

The model uncertainty analysis showed that the model is generally robust to measurement errors, other than when there is an error in either the CER or the OUR data, but not both. This is due to the importance of the respiratory quotient (RQ) in the stoichiometric balance. This should be an area for future work, in order to identify when the OUR or the CER measurement is not in the expected range, so that it is possible to identify when the model prediction may not be valid. In addition, it was shown that the model has an unacceptable uncertainty range when we consider the parameter uncertainty to be from fitting the six stoichiometric parameters simultaneously. However, if one set of parameters is fixed, for the product concentration for example, and the others are fitted, the model parameter uncertainty is less than 5.5%. This is due to the high correlation between the parameters in this stoichiometric balance model.

An area of future work for this method, could also be to consider the impact of product glycosylation patterns on the resulting stoichiometric balance. For fungal systems, it is known that a significant proportion of the product mass is due to glycosylation [Hui et al., 2002, Górka-Nieć et al., 2010]. Therefore it is interesting to investigate the product composition with respect to the level of glycosylation. This is an extension to the uncertainty analysis carried out in this work. Similarly, the biomass composition may also be affected by changing cell wall composition [Górka-Nieć et al., 2010].

A weakness of the model structure is the separation of the substrate concentration from the stoichiometric balance, which results in the possibility of a negative substrate concentration. As discussed, there may be valid explanations for a negative substrate concentration, due to an error in the assumed feed substrate concentration, an incorrect starting substrate concentration from the batch phase, or consumption of other media components which contain carbon source. This however cannot be validated without further data available, so future work includes further investigation of these factors. Another possibility is to limit the consumption of feed in the stoichiometric balance when the substrate concentration is below a threshold level, however this is not desirable as the biomass and product profiles fit the experimental data well, and this would lead to cases of under prediction of these states. The substrate concentration balance is the greatest area of uncertainty in this modelling work and must be further investigated.

With respect to the dissolved oxygen concentration in the system, the model shows significant fluctuations, which are in correlation with the oxygen uptake rate, however this is not reflected in the data. The dissolved oxygen prediction should therefore be further investigated in the model.

In order to apply more advanced optimisation and control strategies to industrial fermentation processes there is a need for robust state estimators in order to identify key performance indicators in real time. The future work for this project is to further develop this monitoring tool, investigate further the model robustness, and work towards on-line control and optimisation at pilot and production scales.

Part IV

Control

7 Control strategy development

The overall aim of this thesis is to develop a control strategy to maximise the product produced from a fed-batch fermentation process. The previous chapters have focussed on the development of modelling tools in order to aid control strategy development. A mechanistic model for off-line application has been developed into an on-line monitoring tool. This allows on-line estimation of parameters which are otherwise not available in real time. This final part of the thesis describes the application of the modelling and monitoring tools in order to improve the process by development of a control strategy. This chapter provides a summary of the control objectives, and describes the control strategy development. This control part of the thesis is based upon the following article:

A novel model-based control strategy for aerobic filamentous fungal fed-batch fermentation processes

Mears L., Stocks S. M., Albaek M. O., Cassells, B., Sin G., Gernaey K. V.
Biotechnology and Bioengineering, Accepted, doi:10.1002/bit.26274 (2017)

7.1 Control objectives and motivation

The aim of this project is to reconsider the control approach for fed-batch fermentation processes. This includes a move away from single-input-single-output (SISO) feedback control systems, and towards a model-based approach, allowing improved control and monitoring capabilities for the process. By utilising models, this allows for a range of control objectives to be defined, as opposed to being limited to set point tracking, as applied in feedback systems. It is therefore possible to define freely the controller objective.

The overall objective for an industrial fed-batch fermentation process is to maximise profit. If the system boundary is considered to be the fermentation operation alone, then the following aspects are to be considered:

- Total product mass per batch
- Value of product
- Cost of power for agitation
- Cost of power for aeration
- Cost of power for air compressor
- Cost of power for reactor cooling
- Cost of raw materials

For high viscosity filamentous fungal processes, oxygen can become the limiting substrate as the process develops [Posch et al., 2013]. Oxygen limited conditions reduce the productivity of the processes, and therefore this should be avoided. For this reason, it is not of interest in this work to try to account for the cost of power to the system inherently in the optimisation problem, as it is known that the maximum vales are required in order to achieve the necessary oxygen transfer rates. In addition, it has been shown that the dominating operating cost for such processes is raw materials, rather than power [Albaek, 2012]. Albaek (2012) shows that with a similar process, the cost of raw materials was 5-7 times higher than the cost of electricity due to aeration, agitation and cooling. In this way the impact on cost is limited. For the purpose of this project, the optimal process is therefore considered to be that which maximises the total product mass per batch, given a scheduled batch time.

This control objective also relates to the continued pressure in industry to maximise available production capacity [Stocks, 2013]. Increasing product demand must be met by an increase in capacity, however investing in new stainless steel vessels is costly. It is important first to make full use of the available capacity. This requires that the maximum product is achieved consistently in every batch.

In order to maximise the product mass, it is possible to either increase the product concentration or to increase the mass in the system. There are many control strategies which specifically aim to increase the product concentration obtained from a fed-batch fermentation process [Chang et al., 2016, Kovárová-Kovar et al., 2000, Peng et al., 2013]. Increasing the product concentration is often focused on reducing by-product formation, which is a waste of substrate, and a burden for the process. However, in industrial fermentation processes, the host strains are highly engineered and optimised, and many of the pathways associated with by-product formation are removed to optimise productivity [Cherry and Fidantsef, 2003]. This is the situation with the industrial strain of interest in this work. For this reason, it is considered that the final mass in the system is the factor which can most greatly impact the total product mass. It is also subject to variation between batches.

7.1.1 Maximising batch fill

Maximising the fill of a fed-batch process is not a trivial problem. The system must be initiated at a level which means the target is reachable, given the oxygen transfer conditions of the system. This is since the oxygen transfer rate in the system at a given time, limits the feed rate which may be applied, in order to maintain aerobic process conditions. This rate is constantly changing due to the dynamic conditions in the batch process; the biomass concentration increases, the viscosity increases, the total mass increases and therefore the power input per volume decreases, and these factors all contribute to a reduction in the oxygen availability. This means that if the start fill of the batch is too low, it will not be possible to achieve the target in a fixed batch time without over feeding the system. This dependence on the rheological properties of the broth, as well as the specific substrate uptake rates, means that the optimal start fill is therefore strain specific.

For a given strain, the start fill is also dependent on the process operating conditions for the batch. This is, for example, with respect to the headspace pressure, the agitation rate, aeration rate, and temperature, all of which affect the rate of oxygen transfer to the broth. This means that the start fill is not only strain specific, but also dependent on the process operating conditions. This issue may be exemplified by analysis of the mass profiles for the eleven batches used for model development, as described by Albaek et al. (2012). In Figure 7.1, the mass profiles for

the eleven batches are shown, when each of the eleven batches had different process operating conditions, corresponding to a full factorial design of two levels for three process parameters; power input, aeration rate and headspace pressure [Albaek, 2012]. This shows that there is a major difference in the feed capacity due to the different operating conditions tested, and that this must be accounted for in the batch planning phase, in order to make full use of the available productive capacity.

The start fill is also dependent on the equipment used. Different tank dimensions, and different impeller types, the number of impellers and configurations will result in a system with different oxygen transfer rates. It is therefore not correct to assume the optimal start fill is equivalent between equipment, without first an assessment of the physical conditions in each system. Similarly, it is not possible to apply the same relative start fill to two different scales of tank. This is related to the well documented issues of process scale up, as described in detail in Stocks (2013). If a start fill is defined for a pilot scale process, the same relative fill is not directly applicable to a larger scale process, as the physical conditions do not scale linearly with size. For example, the additional hydrostatic pressure in a large process increases the oxygen solubility, and improves oxygen transfer. The Reynolds number is also greater at scale, due to the proportionality to the square of the agitator diameter, resulting in turbulent mixing at lower stirrer speeds [Stocks, 2013]. Overall, it is clear that the start fill is dependent on the strain, the process conditions, the scale of operation, and the equipment used.

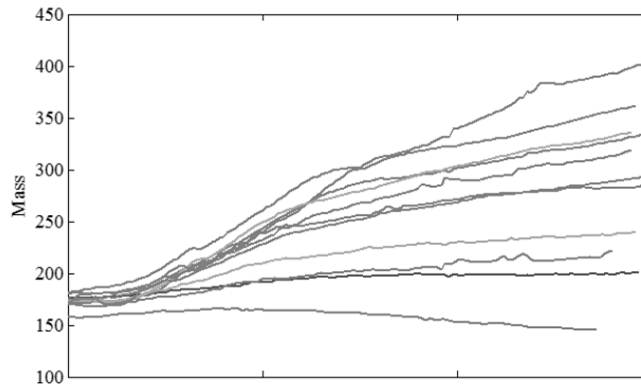


Figure 7.1: Mass profiles for the eleven batches of data used for mechanistic model development as described in Albaek et al. (2012), employing different process operating conditions for headspace pressure, aeration rate and stirrer speed.

Assuming an appropriate start fill is identified for a fermentation process, it is also not possible to guarantee that the target fill is reliably achieved in the system. This is due to many batch-to-batch uncertainties which may be experienced:

- The batch phase of a fed-batch process may vary in length due to different initial biomass concentration, age of the inoculum culture, or other batch-to-batch variations. This means that the feeding time may not be equal in every batch.
- The feed concentration is subject to batch-to-batch variation, meaning the total mass added for a given mass of substrate is not equal in every batch.

- Physiological parameters, for example mean growth rates, yields and maintenance requirements are slightly different between batches, resulting in different metabolic rates.
- Evaporation rates are dependent on the outside air conditions and are therefore not constant between batches.

Therefore, for a given process, it is expected that even with the same start fill, there will be a considerable variation in the final fill. This variation may be reduced by means of a control strategy which aims to target a maximum batch fill.

The concept of this work is to reduce the variance in the final batch fill and consistently achieve maximum tank capacity. This control objective has been defined in order to maximise the product mass in a batch operation, however this objective also brings additional benefits when considering the full industrial processes. The fermentation process is just one step in the production train, which includes product recovery, and formulation. In batch process operations, it is challenging to schedule operations in order to meet a certain demand due to uncertainties in the batch which can have a large impact on the full downstream process. In this case, having a very well defined final product mass from the upstream process can aid planning and scheduling operations downstream.

7.2 Controller development

The aim is to reduce the variance in final mass, and target the maximum mass in the process, however, if we reduce the variance in one process parameter (tank fill) there must then be variation in another processing parameter. In order to assess this trade off, it is possible to analyse the relationships between the process variables. Figure 7.2 shows a representation of the system being considered. Focussing on the mass in the system as the control target, there are two major influencing variables, namely the feed rate and the evaporation rate. In this work, it is considered that a set of operating conditions are provided for a given process, meaning that the stirrer speed, aeration rate and headspace pressure are defined. This means that the feed rate is the variable which may be adapted in order to reach the desired mass target.

The feed rate is also highly influential in the process, affecting both the mass dynamics, and the oxygen dynamics through the metabolic processes. It is seen from Figure 7.2, that the feed flow rate affects the substrate concentration in the system, which then affects the metabolic rates in the system. One of these rates is the oxygen uptake rate, which directly affects the dissolved oxygen concentration. In addition, the biomass concentration is affected, which causes an increase in the viscosity of the system. This then affects the oxygen transfer into the liquid, and therefore also affects the dissolved oxygen concentration. This shows how the feed rate is such a highly influencing variable to a fed-batch process, affecting the mass dynamics and the metabolic processes and therefore the oxygen dynamics.

The controller objective is to target the final mass, whilst avoiding oxygen limitation in the system. Therefore the feed rate manipulation is considered a trade off between achieving the final fill, and avoiding oxygen limitation. By use of a predictive model, it may be possible to guide the process to the target mass whilst also accounting for the oxygen dynamics.

This strategy assumes that the process operating conditions are fixed. In this work, the process will have a fixed maximum level for the operating conditions, however it is possible to operate under this maximum value if it is desirable. Usually this would not be desirable, as this reduces the oxygen transfer to the system. However, it is possible to manipulate the evaporation rate by

use of the process operating conditions for the headspace pressure and the aeration rate. It is therefore conceivable that these parameters may be down regulated in response to fill. However at this stage in controller development, the focus is on manipulating the feed rate.

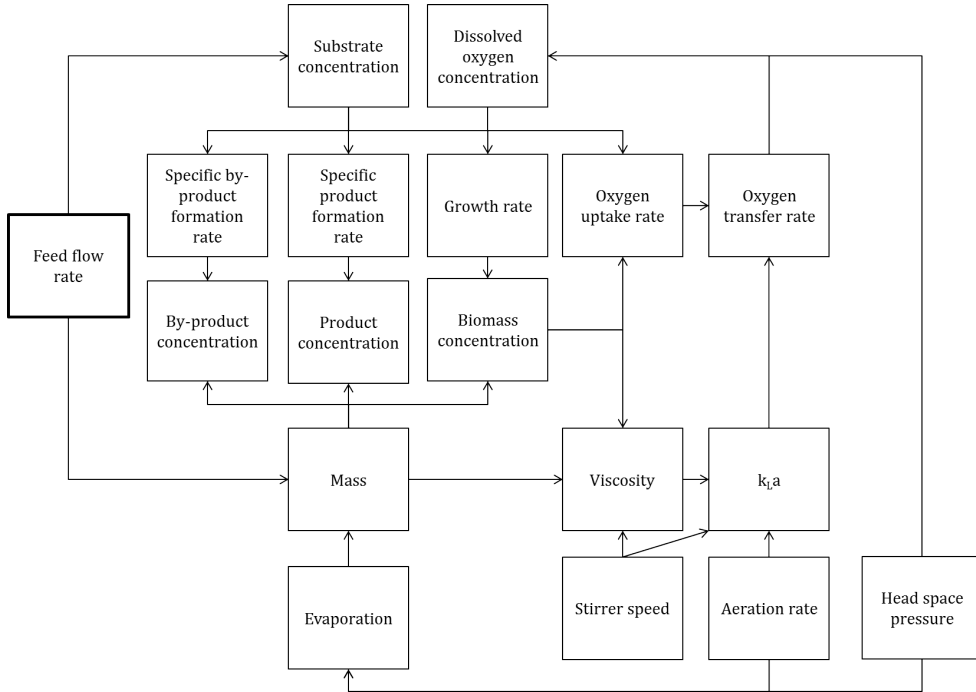


Figure 7.2: Process diagram showing the feed flow rate effects on the process. In order to control the mass, the feed flow rate is manipulated, when the aeration rate and the headspace pressure are fixed for a given process.

In this work a control strategy is described for controlling a fed-batch process to achieve maximum fill capacity subject to the oxygen transfer limitations of the system. This also includes a start fill calculation which is found by use of a mechanistic process model. There is no known reference to such a control objective described in the literature, so this proposes a novel approach.

8 Model-based batch planning

8.1 Introduction

In Chapter 7, the problem statement for the control strategy is described in detail. The control objective is to target a maximum tank fill, reproducibly, whilst avoiding oxygen limitations. This approach requires a batch planning phase where the start fill for the batch is defined, followed by an on-line control strategy. This chapter describes the methods and results of the batch planning phase, and the following chapter then discusses the application of the planned start fills with the on-line control algorithm. The separation is made clear as the planning is an off-line task which applies the model in a form which is suitable for off/line application. The control algorithm is run on-line and also uses the model in the form for on-line application.

In this chapter the mechanistic model is solved in order to define appropriate start fills depending on the process conditions. The model is challenged with four different sets of operating conditions. This is a test of the robustness of the method. It is difficult to make conclusions on the suitability of the start fill if it is run in collaboration with the on-line control strategy, as it is the purpose of the controller to target the maximum fill and reduce variance in this way. Instead the start fill is tested with a reference control operation; namely dissolved oxygen control using the feed rate as a manipulated variable using a PI controller. A full description of the methods and the results for the model-based batch planning is provided.

8.2 Methods

8.2.1 Mechanistic process model

The mechanistic model applied for the purpose of batch planning, is not the same as the model developed in Chapter 6, as the model must be applied off-line. The model structure is the same, and the mass basis is used, rather than a volume basis as discussed in Chapter 5, however instead of the parameter estimation, the yields and maintenance terms are used as described in the original model description, described in detail in Section 5.2.1. This is since there is no available measurement data when applied off-line in order to make the parameter estimates.

The necessary process variables to define for the model are the maximum operating conditions for the headspace pressure, aeration rate and the stirrer speed. In addition it is necessary to define the relative humidity and temperature of the outside air. This must be estimated, and was defined as a constant for each of the batch times, although in reality it is known that these parameters are time dependent. The operating conditions used are described in the next section.

8.2.2 Process operation

The process was run in Novozymes pilot plant facility using a proprietary filamentous fungus production strain in 550L stirred tank bioreactors. Frozen spores were seeded onto agar plates, before inoculating a seed tank. The seed tank conditions are proprietary. In order to ensure the same initial conditions for the batches, despite different start fills, the main fermentation tanks were inoculated from this seed tank with approximately 10% volume. The calculated start fill for the main tanks was then made up of 90% media and 10% inoculum resulting in the same biomass concentration in the initial batch phase. The batch phase for all four tanks have the same operating conditions for stirrer speed, aeration rate and headspace pressure, where the minimum set points are applied (- - - as referenced in Table 8.1). No absolute values for the conditions are provided in this work due to confidentiality.

The feed was started once the batch phase carbon source was depleted. Feed addition was implemented using a dissolved oxygen PI controller. At the point of feed start, the process conditions were then changed from the minimal values to the set points shown in Table 8.2, over a fixed number of hours, in order to make a gradual transition to the new set points. Maximum feed rate constraints were also applied.

The maximum tank fill was defined as 410 kg, and the total batch time was fixed for all tanks and all experiments. The total batch time consists of a batch phase of operation, followed by a feeding phase. In order to estimate the expected feeding phase length from the total process time, past batches of data are analysed to determine the average batch phase duration.

8.2.3 Process conditions

Four sets of process operating conditions were used in this work in order to challenge the methods developed. These are shown in Table 8.1. The exact values of the processing conditions are not provided for confidentiality reasons. In an industrial setting, different production sites may have different constraints on the operating conditions. This is due to different ages and design of equipment in different locations. This is another rationale for the use of different operating conditions, as in an industrial context it is desirable to account for these differences in the planning phase. For consistency, the different process conditions are always referred to as tank numbers 1-4, as provided in Table 8.1.

Table 8.1: Experimental design showing the four different process conditions applied. Absolute values of the process conditions are not provided for confidentiality reasons. Tank numbers are referenced throughout this work as shown in this table, and plot colours are also consistent.

Plot colour	Tank	Stirrer speed	Aeration rate	Pressure
	1	+	+	+
	2	+	-	-
	3	-	+	-
	4	-	-	+

8.2.4 Batch planning methods

Two different methods were applied in order to solve the desired start fill; optimisation of the feed rate only, and optimisation of the feed rate and the operating conditions.

Feed rate optimisation

A non-linear optimisation problem is defined where the model is solved every hour in order to follow a predefined dissolved oxygen set point, by manipulation of the feed rate to the system. The unconstrained non-linear optimisation problem was solved using MATLAB [The MathWorks Inc., 2013], and applying the *fminsearch* algorithm. A reference dissolved oxygen profile was used in order to solve the feed rate addition. This profile may be seen in Figure 8.1. This also requires an estimate of the batch phase duration, prior to feed start. This was approximated by analysis of past batch data for this strain.

$$\underset{F}{\text{minimize}} \quad (DO(F) - DO_{set})^2 \quad (8.1)$$

Simultaneous feed rate and operating condition optimisation

An alternative batch planning tool was developed, where the feed rate is solved in addition to the set point for the operating condition at each time point. This is in order to allow for the set points to increase only as necessary based on the current oxygen transfer limitations. It also avoids issues of overfeeding in the initial batch time. This was done by a constrained optimisation procedure using the *fmincon* algorithm, also implemented in MATLAB [The MathWorks Inc., 2013].

$$\begin{aligned} &\underset{U}{\text{minimize}} \quad -(P(U)M(U)) \\ &\text{subject to} \quad U_{min} \leq U \leq U_{max}, \quad f(U) \leq 0 \end{aligned} \quad (8.2)$$

In this case, U is defined as the inputs to the process, namely the feed rate, stirrer speed, headspace pressure, and aeration rate (F , N , P and Q_{air}). These four inputs were solved simultaneously, subject to constraints. Constraints were applied to the minimum and maximum dissolved oxygen concentration, as 20% and 100% respectively. The maximum mass in the system was 410 kg, and the maximum operating conditions were as defined in Table 8.1.

8.2.5 On-line measurements

The start fills were applied to two reference experiments, and in this case on-line data was measured from the processes. On-line data was logged by a data acquisition system. Dissolved oxygen concentration was measured on-line, where the probes were calibrated to the same % saturation at the same processing conditions prior to the process. Other on-line measured variables include the mass, feed flow rate and base flow rate, as well as the operating conditions for the processes. The oxygen uptake rate was calculated on-line, based on the oxygen difference in the oxygen partial pressure in the inlet and outlet gas flows, and knowing the air flow rate. The carbon dioxide evolution rate was calculated in the same manner. An additional on-line measurement was the on-line viscosity, which was measured using a Hydramotion XL7A probe.

8.2.6 Off-line sample analysis

The biomass concentration was measured by dry mass determination, by drying at 105 °C for 48 hours. The biomass sample was washed twice with deionised water to remove soluble media components. The product concentration was determined based on a generic protein assay used at Novozymes A/S. Viscosity was measured off-line in an AR-G2 rheometer from TA instruments using a vane-and-cup geometry. The vane consists of four blades at right angles (14 mm x 42 mm), the cup had a 15 mm radius, and the gap between the vane and cup was 4000 μm . Measurements were taken in the interval of 10 to 600 1/s and the Bingham plastic model was applied to describe the rheological behaviour [Bingham, 1916]. The shear rate for apparent viscosity determination was found by the approach of Metzner and Otto, ksN , where ks is 11.

8.3 Results and discussion

8.3.1 Start fill determination

The start fill was calculated for the four different sets of operating conditions as shown in Table 8.1. For each tank, the full process is simulated based on the desired dissolved oxygen profile, and the defined process operating conditions, as well as fixed parameters for the outside air temperature and relative humidity, which along with the operating conditions, define the evaporation rate from the system. Figure 8.1 shows an example simulation of the full process time for Tank 4, showing the start fill, in order to reach the desired maximum fill in a defined process time, as shown in Figure 8.1 by the asterisk.

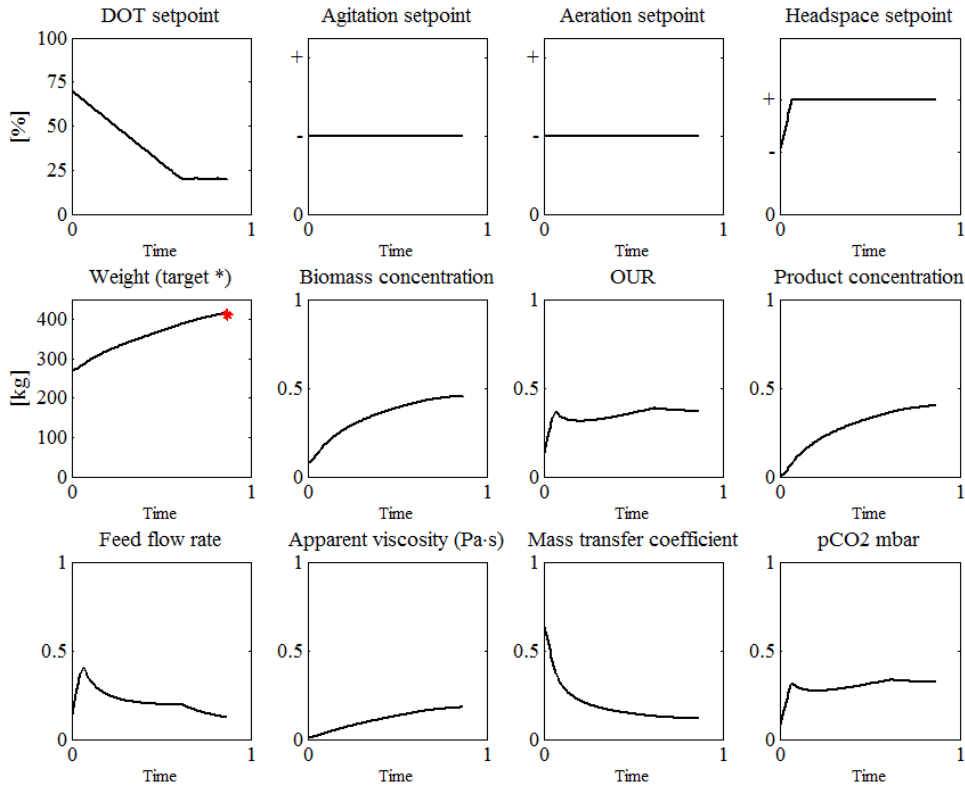


Figure 8.1: Example of the results of batch planning for Tank 4. Provided with the process operating conditions, and a desired dissolved oxygen profile, the model is solved by manipulating the feed rate at each hour. This results in a mass trajectory which is compared to the desired mass target in order to identify an appropriate starting mass. Where axis labels are missing it is for confidentiality reasons. Note, the batch-phase of the process is not included, only the fed-phase is shown.

The full process simulation shows how the dissolved oxygen set point is achieved by manipulating the feed rate, for given process operating conditions. The dissolved oxygen profile is shown, as well as the ramp in the headspace pressure, which is predefined, in order to reach the desired operating condition in a fixed time after feeding is initiated.

The same result is shown for Tank 1 in Figure 8.2, which is that with maximum process conditions for agitation, aeration and headspace pressure (+ + +). In this case it is necessary to be critical of the result. This is due to the very high oxygen transfer rate obtained in the initial hours of the fed-phase, where there is limited biomass for oxygen uptake, but increasing oxygen transfer due to the increasing set points for the process conditions. This leads to an excessively high feed rate in order to solve the model for the desired dissolved oxygen profile. In the real process operation this effect is also experienced when using a dissolved oxygen feedback controller, as there can be overfeeding in the initial hours of the batch, where the dissolved oxygen concentration is above set point, but the limiting factor is not the availability of substrate, but instead the concentration of biomass for oxidative metabolism of the substrate which is present.

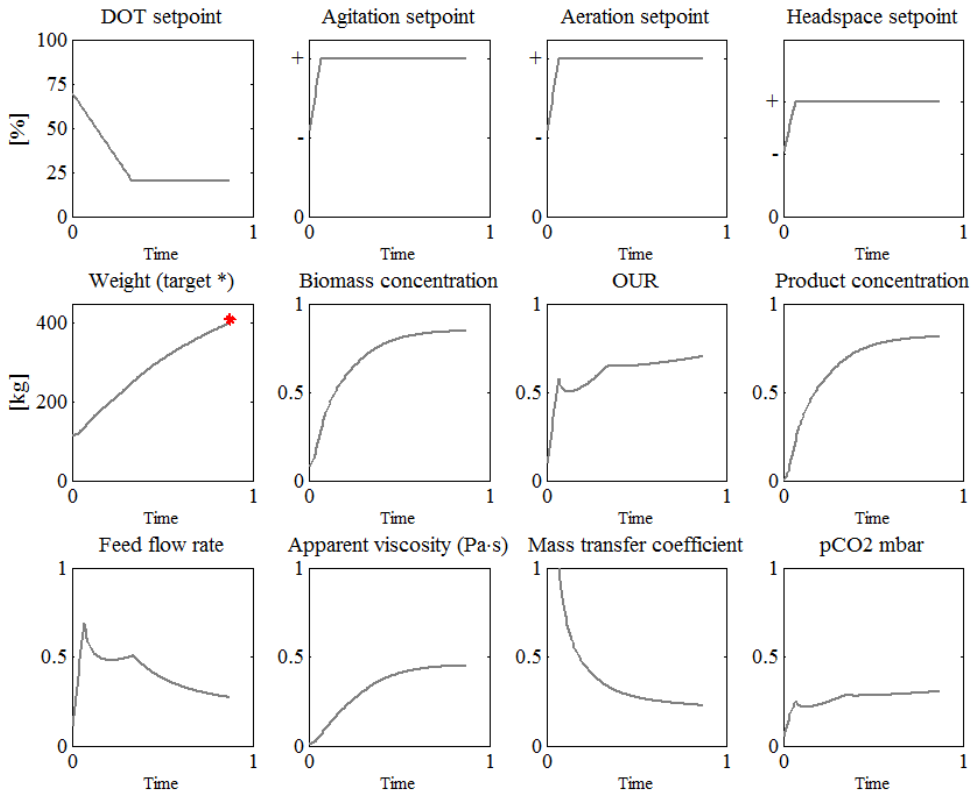


Figure 8.2: Results of batch planning for Tank 1, where the process operating conditions are defined, and the feed rate is solved every hour in order to obtain a desired dissolved oxygen profile. This results in a mass trajectory which is compared to the desired mass target in order to identify an appropriate starting mass. Where axis labels are missing it is for confidentiality reasons. Note, the batch-phase of the process is not included, only the fed-phase is shown.

In order to obtain an estimate of an appropriate start fill in this case it is also possible to allow manipulation of the operating conditions, such that the oxygen transfer rate is only increased as necessary. This is implemented as a non-linear optimisation problem with input constraints and non-linear model constraints. Constraints are provided for the process condition as defined

in Table 8.1, as well as a minimum dissolved oxygen concentration of 20% saturation, in order to be similar to the dissolved oxygen profile. This is shown for Tank 1 in Figure 8.3. The set points for all three of the process parameters is delayed, and increases more gradually than when the fixed ramp is applied. The feed flow rate follows the maximum ramp until the point where the process conditions are at their maximum values, and the oxygen becomes limiting, and then the feed rate is reduced to maintain aerobic conditions. It is also seen that the agitation is reduced once the maximum weight is reached, in order to reduce the feed rate applied. Also, the headspace pressure is reduced in order to increase the evaporation rate in the system, and maintain the fill despite the feed being added.

This provides a better estimate of an appropriate start fill, as it does not allow for over feeding in the beginning of the feeding time. It could have been possible to solve the dissolved oxygen profile, as described in Section 8.2.4, but also applying constraints on the feed rate only, however this led to model instabilities, as it is not possible to achieve the desired dissolved oxygen profile, with a limited feed addition. It was for this reason that the other process parameters were also solved simultaneously. For the systems with high headspace pressure set point this comparison between the two methods has been made in order to obtain an appropriate start fill estimate. The results are not greatly different, as the issue only occurs for a short part of the process time, however it is interesting from a perspective of understanding the process operation, and identifying opportunities for future improvements, for example in the set point profiles for the process variables.

Table 8.2 shows the results of the start fill calculation which was obtained by simulating the full process using the four different sets of operating conditions. It is seen that the difference in the stirrer speed has the greatest impact on the start fill, as it has the greatest impact on the oxygen transfer. Of course, without absolute values of the process parameters, this is of limited value to the reader, however it will be important to identify if this trend is also represented in the experimental data. For example, if the lowest start fill is also the lowest end fill, then it is likely that the start fill was not appropriately determined. The start fill ranges from 39% to 77% of the target fill depending on the processing conditions. In order to validate if this difference is of correct magnitude, it is necessary to run experiments utilising the different start fills. It is not desirable to run the model-based controller for this purpose, as we wish to identify the variation in final fill for the given start fills. Therefore dissolved oxygen control is applied, as a reference PI controller, using the feed rate as manipulated variable.

Table 8.2: Start fill calculation using mechanistic process model for different processing conditions. Process conditions are not shown for confidentiality reasons. Start fill shown as a percentage of the target fill.

Plot colour	Tank	Stirrer speed	Aeration rate	Pressure	Start fill
	1	+	+	+	39%
	2	+	-	-	49%
	3	-	+	-	77%
	4	-	-	+	67%

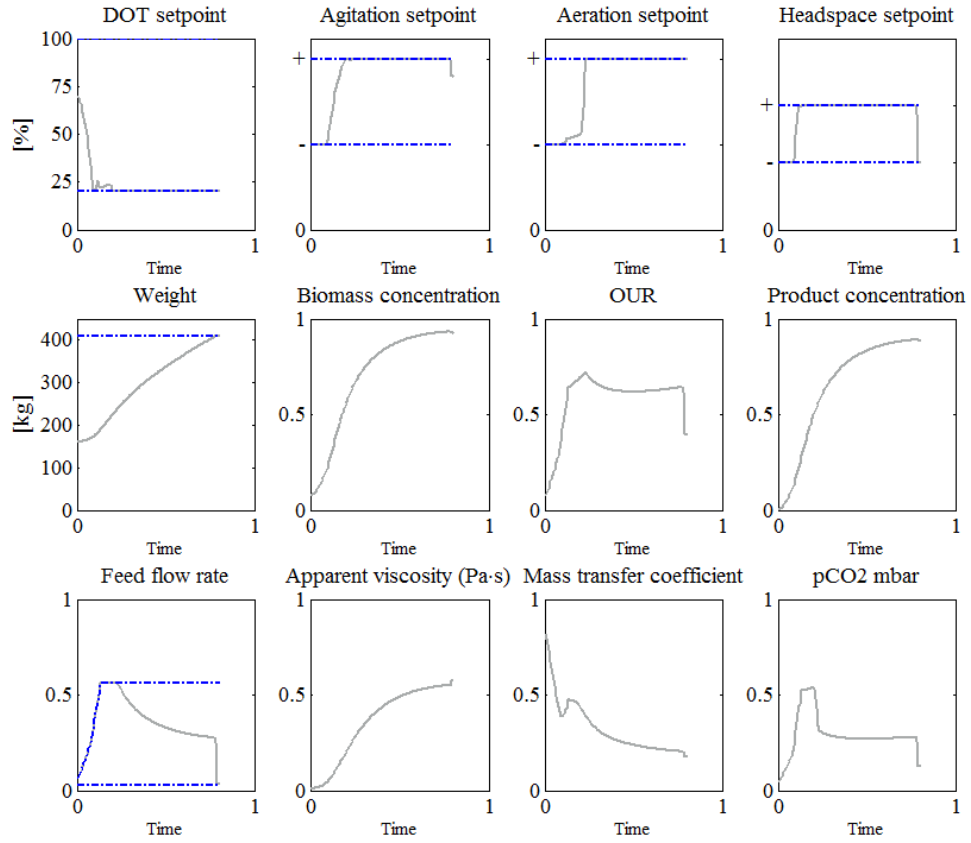


Figure 8.3: Results of batch planning for Tank 1, where the feed rate, agitation rate, aeration rate and headspace pressure are solved simultaneously, subject to input constraints (blue dashed lines) and the non-linear model constraint in order to maximise total product at each time point. This results in a mass trajectory which is compared to the desired mass target in order to identify an appropriate starting mass. Where axis labels are missing it is for confidentiality reasons. Note, the data shows only the phase of operation when feeding has been initiated.

8.3.2 Dissolved oxygen controlled experiments

Final fill

The results for the two DO controlled experiments (Experiments 1 and 2) are shown in Figure 8.4. The start fill and final fill is shown for the two experiments, each run with the four different operating conditions, as described in Table 8.2. It is seen that the final fill of Experiment 2 is noticeably lower for all conditions than Experiment 1, despite the very similar start fills. This emphasises the relevance of the evaporation rate, in long running aerated processes as described in Chapter 6.

Due to the high fill levels seen in Experiment 1, some manual interventions were made to ensure that the tanks would not in fact overflow, which is a serious problem in a production facility where biological material must be contained at all times. For this reason, tanks 2 and 3 had small reductions ($<25\%$ change) in their stirrer speeds in experiment 1, since they reached the point where they were 10% over the target fill. This was done as a safety precaution. This also highlights the need for control of the fill in a tank, as it can be hazardous to allow the tank to become over full.

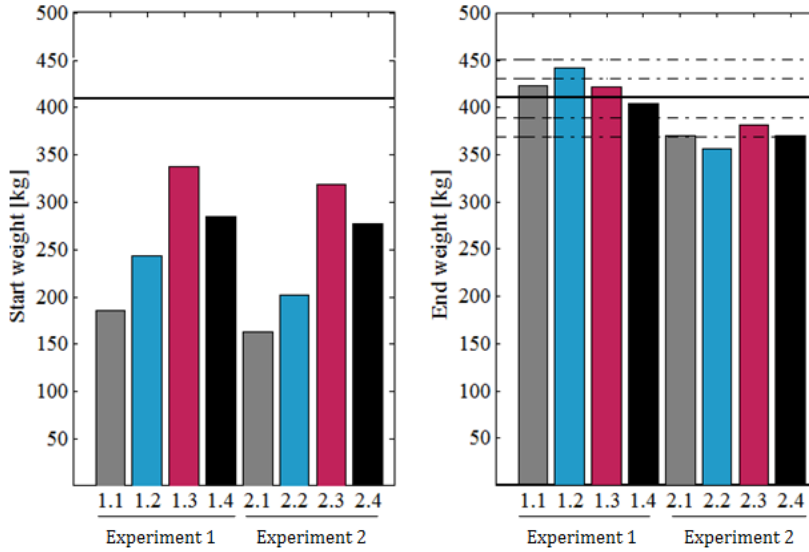


Figure 8.4: Start mass and end mass for experiments 1 and 2 and tanks 1-4. Solid line shows the target fill, and dashed lines show 5% and 10% deviation from the target end weight

Despite the need to intervene, it is considered that the start fills are of a suitable magnitude, and the model captures the different oxygen transfer rates due to the process operating conditions with a suitable accuracy. Focusing on Experiment 2, where no changes were made to the operating conditions specified, the difference in start fill of 156 kg was reduced to a difference of only 24 kg in the final fill. This large reduction in variance shows that there is indeed a robust process model which is applicable to defining an appropriate start fill given a set of operating conditions. However, these two experiments have also emphasised that the evaporation rate is of significance and should not be considered constant between experiments. For this reason also, the start fill is not fixed at an optimum, but will be dependent on the relative humidity and temperature of the outside air.

Evaporation

The two experiments were operated three months apart, and so it would be expected that they are subject to different air conditions. In order to assess the difference in the evaporation rates at the two experiments, it is possible to look at the batch phase of the experiment, and see the difference in mass change over this period of the operation. In this phase there is no mass addition as no feed is added, and all experiments have the same operating conditions; it is only after the feed start that the process conditions are adjusted for the different tank numbers. Figure 8.5 shows the batch phase mass change for the two experiments and for all batches. Although it is acknowledged that there is significant variation between the tanks in a single experiment, it is seen that Experiment 2 has a greater mass change than Experiment 1 for all tanks. This can partly explain the deviation in final mass between the experiments in Figure 8.4.

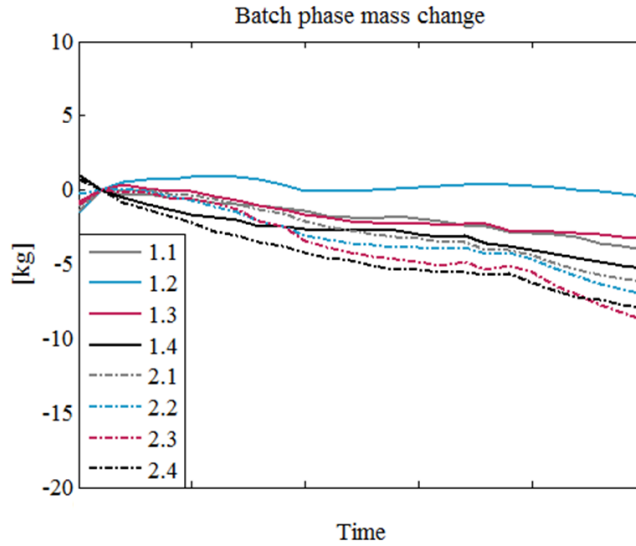


Figure 8.5: Batch phase mass change (kg) for experiment 1 (solid lines) and Experiment 2 (dashed lines). In the batch phase of the operation all the tanks have the same operating conditions and no mass is added, so this mass change is due to evaporation. Time not shown for confidentiality reasons.

In addition to looking at the data, it is also important to ensure that the model is able to capture this difference. When the actual air temperature and humidity was used as input to the evaporation model, the average rate of evaporation from Experiment 2 is around 10% greater than Experiment 3 due to the air conditions, so the model is seen to reproduce the trend in the data.

Viscosity

For filamentous fungal fermentation systems the viscosity of the broth is a very interesting process parameter to monitor. The viscosity is typically high for these processes due to the filamentous morphology of the biomass. This high viscosity can be problematic from a processing perspective, as it contributes to the oxygen transfer limitations experienced in such processes, and this is discussed in more detail in Section 7.1.1. In addition, the fluid shows non-Newtonian behaviour, meaning that the viscosity is shear rate dependent, and this causes challenges for

the measurement.

The on-line viscosity has been measured using an in-line viscometer, in addition to the off-line sample analysis. It is clear from Figure 8.6 that it is hard to relate these two measured parameters. The viscosity measurement which is commonly used for describing the fermentation broth is the apparent viscosity, which is the viscosity at a shear rate which is selected to represent the average conditions in the tank at a given stirrer speed. The on-line viscosity measurement however, is the value measured at very high shear rate, and due to the non-Newtonian property of the fluid, is of a different scale to that which is measured off-line. This is not a disadvantage, however it makes it challenging to interpret the data.

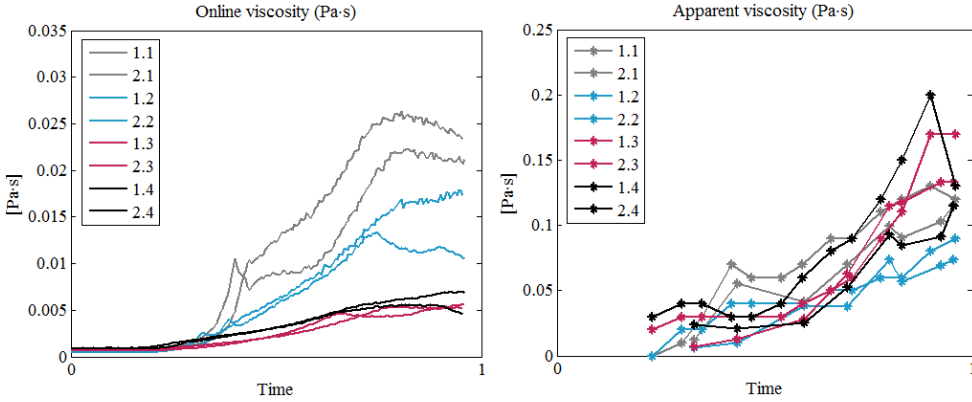


Figure 8.6: On-line and off-line viscosity comparison, for the dissolved oxygen controlled experiments. Absolute values of time not shown for confidentiality reasons.

It is seen from the off-line viscosity data that there is a sharp increase in the viscosity at the end of the fermentation for Experiment 2, Tanks 3 and 4 (2.3 and 2.4 in Figure 8.6). For filamentous fungi, the viscosity is not only dependent of the biomass concentration, but also on the morphology of the biomass. It is expected that this rapid increase is a result of morphological or structural changes in the biomass, rather than the amount of biomass. This cannot be modelled with the current empirical formula used. It is beyond the scope of this work to try to model morphological changes in the biomass, however it is important to realise the model limitations, as this will impact the performance of the model prediction. If the viscosity in the system is underestimated, then the oxygen transfer rate is over predicted and this may lead to issues with the controller algorithm if not identified.

The dataset used to develop the original mechanistic model did not show this form of trend. This suggests that this is a characteristic of the strain applied in this work, which results in this behaviour. The hypothesis is that this is due to morphology, but it is difficult to confirm without dedicated studies. The result of this however, is that the viscosity prediction for the strain used in this work is very poor for the conditions of low oxygen transfer, Tanks 3 and 4, as shown in Figure 8.7. The prediction is acceptable for tanks 1 and 2, but is considered unacceptable for Tanks 3 and 4, where this sharp increase in the viscosity is not captured. It is, however, beyond the scope of this work to incorporate morphological changes into the model.

Due to the limitations in the viscosity model, it is interesting to investigate the on-line viscosity measurements further, as it is very valuable for monitoring purposes if it is indeed a robust

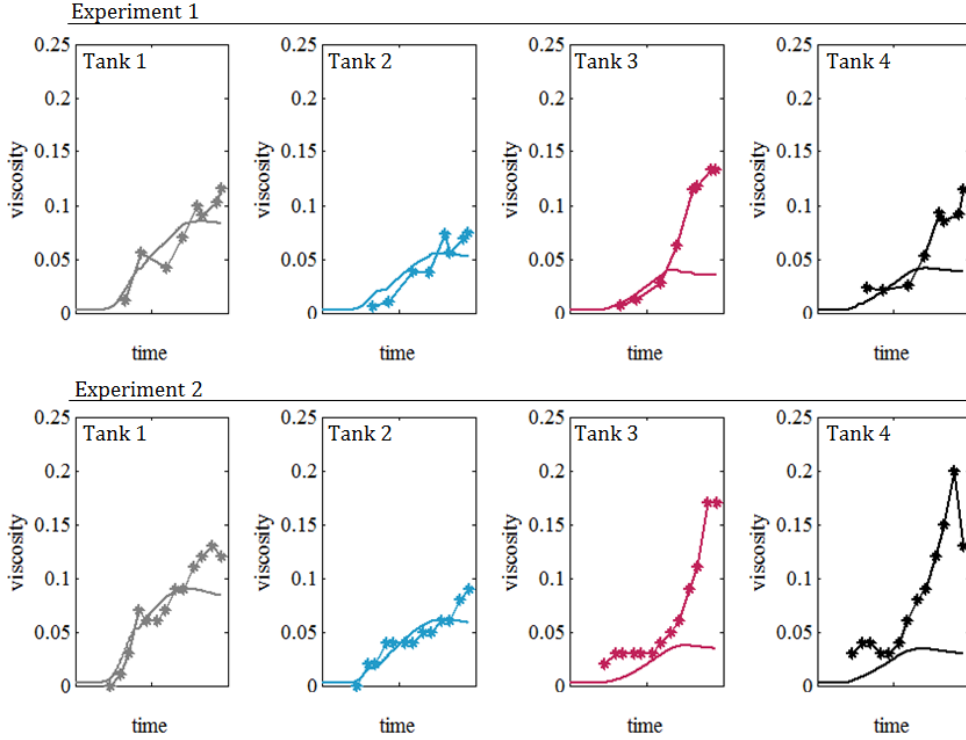


Figure 8.7: Dissolved oxygen controlled experiments apparent viscosity data and model prediction, shown for Experiment 1 (top row) and Experiment 2 (bottom row). The results show that the simple empirical correlation describes Tanks 1 and 2 to an acceptable accuracy, however in Tanks 3 and 4 the prediction is very poor.

measure of the viscosity of the system. Viscosity is currently modelled in order to understand the oxygen mass transfer rate in the system, k_La . Figure 8.8 shows the log-log plot of the measured k_La obtained by the direct method, and the measured on-line viscosity. There is a strong correlation meaning that the k_La is related to the on-line measured viscosity to the power of the slope of the curve, in this case -0.8. This is in the form of the current empirical relation used to estimate the k_La , in the system:

$$k_La = a \frac{P_{total}^b}{V} v g^c \text{viscosity}^d \quad (8.3)$$

The correlation seems to hold very well, other than for Tanks 3 and 4 at the end of the experiments, when the k_La falls rapidly, and this is not captured in the on-line viscosity. This is the period of the batch where the off-line viscosity shows a sharp increase. This suggests that the on-line viscosity measurement also does not capture the phenomenon which causes this. As previously described, it is proposed that this is due to morphological changes. This is an area which should be further investigated in relation in order to better understand the strain used in this work. Additional discussion of the viscosity measurements for the reference experiments is provided in the Appendix.

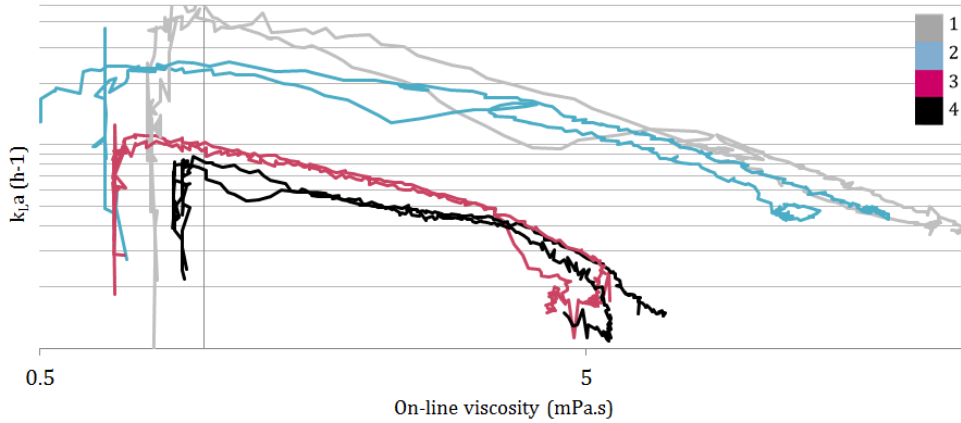


Figure 8.8: Log-log plot of k_La vs on-line viscosity for the dissolved oxygen controlled batches. A linear correlation is shown, with a slope of -0.8.

Dissolved oxygen concentration

The dissolved oxygen concentration is measured on-line in all batches, and in these experiments it has been controlled to a predefined set-point ramp using the feed rate as the manipulated variable. Figure 8.9 shows the dissolved oxygen profiles for the four tanks and the two experiments.

It is seen from Figure 8.9 that there is a trend of over feeding in the initial stages of the fermentation, which is seen by an under shoot in the dissolved oxygen concentration. This is most pronounced in Tank 1, and also significant in Tank 4. This is due to the headspace pressure increasing to the operating condition for the batch, following a ramp over a fixed number of hours. The headspace pressure has a direct impact on the oxygen solubility and therefore the dissolved oxygen concentration rises by up to 20% over a few hours. This causes some issues for the dissolved oxygen controller which over feeds the system, resulting then in a drop below set point before the stable operation is reached. This is since the initial hours of the fed phase of the batch are not oxygen limited but limited by the biomass concentration.

Another observation from the DO controlled batches is that the dissolved oxygen is not able to be controlled for the full process time in the experiments with poor oxygen transfer. It is seen in Figure 8.9, that in the final stages of the batch, for Tanks 3 and 4, the oxygen transfer is insufficient to maintain the DO at the required level and oxygen limitation occurs. In Experiment 1 the minimum feed rate was not zero but instead a minimal feed rate value, so in Experiment 2 the minimum feed rate was changed to zero to see if the dissolved oxygen could then be controlled to set point, but the same trend is seen. This is a result of the extreme conditions which are being tested in Tanks 3 and 4. The low stirrer speed, in addition to a low headspace pressure for tank 3, and a low aeration rate for Tank 4 results in conditions where the oxygen transfer is too low at the end of the batch where the viscosity is high and therefore the oxygen concentration falls rapidly. This is not desirable, however this is also the purpose of running the DO batches as a reference case, so that it is possible to identify limitations in the operation. It is therefore also not expected that any other control operation should be able to maintain aerobic conditions at this late stage in the process for Tanks 3 and 4.

The reason for the high dissolved oxygen concentration in Tank 1 Experiment 1, in the middle of the fermentation time was due to a maximum feed rate being too low in order to reach the set point. This was corrected once it was identified, and the process was able to return to set point in a very short time.

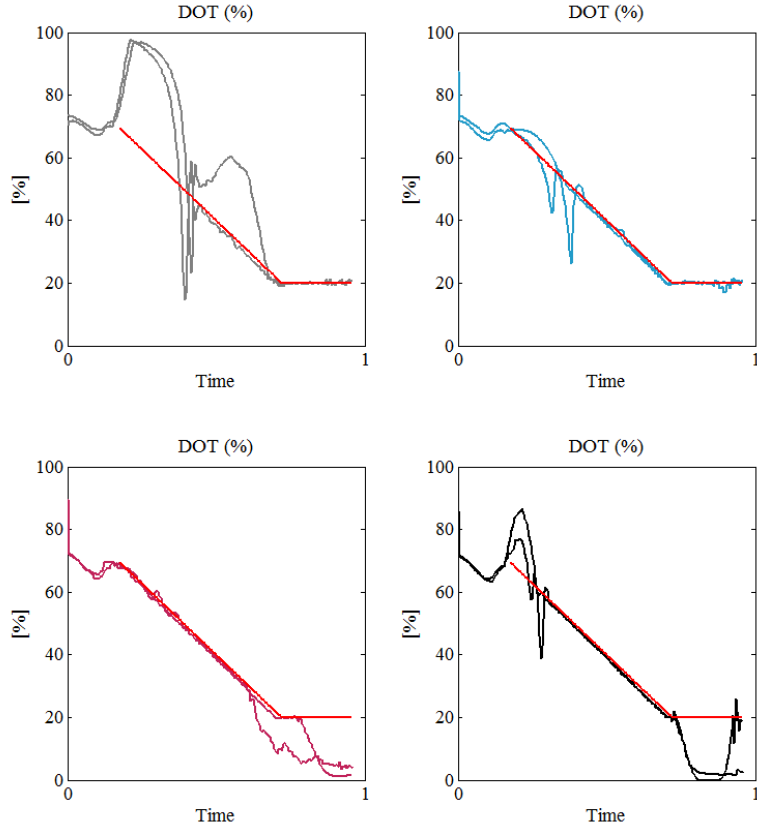


Figure 8.9: Dissolved oxygen profile for experiment 1 and 2. Red line shows the dissolved oxygen set point. Tanks 3 and 4 have poor oxygen transfer conditions (shown in Table 8.1), which is shown to cause oxygen limitation in the final hours of the fermentation despite the use of dissolved oxygen control. Time not shown for confidentiality reasons.

8.3.3 Comparison of batch planning results and experimental data

In order to critically evaluate the batch planning process it is possible to compare the off-line batch planning results to the measured data from the two experiments utilising the calculated start fills and the dissolved oxygen controller. The data is provided for the dissolved oxygen concentration (Figure 8.9), the oxygen uptake rate (Figure 8.10) and the feed rate (Figure 8.11), as it is mostly the physical oxygen transfer dynamics which are tested in the batch planning process.

The results show that the model is able to capture, to an acceptable accuracy, the different oxygen transfer rates in the system, based on the different process operating conditions, as

shown in Table 8.1. It has already been discussed, that the dissolved oxygen concentration becomes limiting at the end of the batch for the tanks with a low stirrer speed (3 and 4), and this is not predicted in the model. Despite this, the calculated feed rate is of correct magnitude for the four different tanks, and this shows that the model is of suitable accuracy for application as a batch planning tool, at different process conditions.

When analysing the results for the feed rate in Figure 8.11 it is also important to remember that the model is solving the feed rate at each time point, but this does not include controller dynamics which are present in the on-line system. This was done as the model will also be used in a model-based control strategy and it is therefore not desirable to include feedback controller dynamics at the batch planning stage. The measured data includes the dynamics of the feedback controller, where it has also been discussed that there is overfeeding in the initial experiment time. It is for this reason that there is a fluctuating feed rate in the data shown. It does appear, however, that the predicted feed rate in the final hours of the batch is too high, not only for Tanks 3 and 4 but for all tanks. This suggests that the oxygen transfer rate falls more greatly than is predicted in the model. This has already been discussed in relation to the viscosity model.

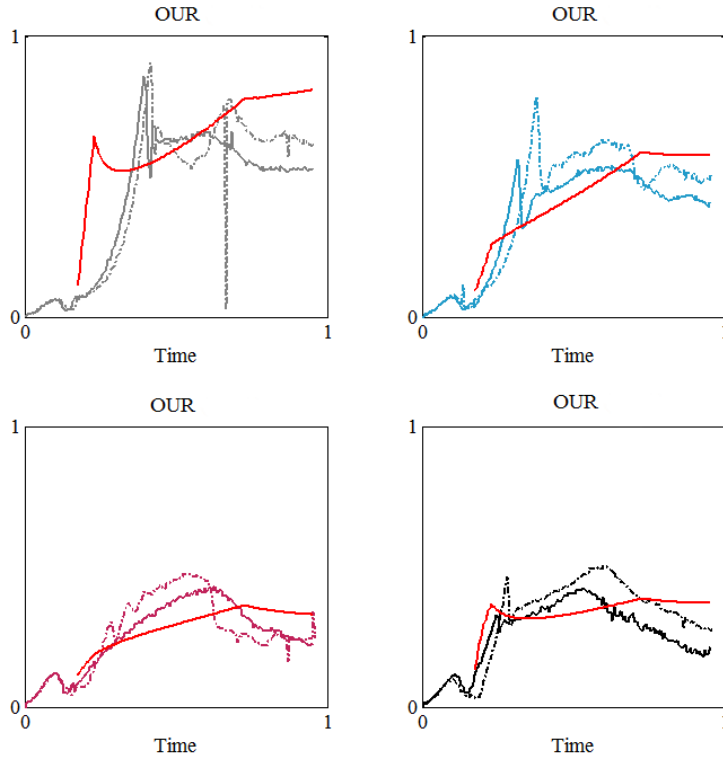


Figure 8.10: Oxygen uptake rate data from Experiment 1 and 2 (colours and tank conditions as shown in Table 8.1), with the corresponding prediction from the batch planning phase (red). Axis values not shown for confidentiality reasons.

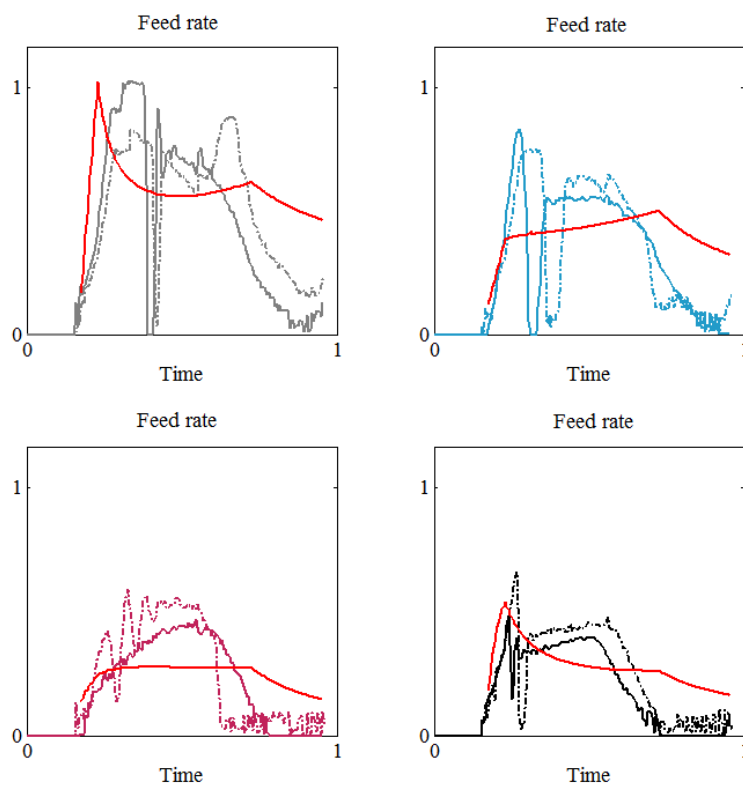


Figure 8.11: Feed rate data from Experiment 1 and 2 (colours and tank conditions as shown in Table 8.1), with the corresponding prediction from the batch planning phase (red). Axis values not shown for confidentiality reasons.

8.4 Conclusions and future work

The start fill calculation has been successfully applied on-line at 550L scale utilising four different sets of process operating conditions and two reference experiments. The results from Experiment 1 and Experiment 2 show that the calculated start fills are of suitable magnitude given the different operating conditions. In Experiment 2, it is shown that the start fill variation of 156 kg was reduced to a difference of only 24 kg in final fill. This large reduction in variance shows that there is indeed a robust process model which is applicable to defining an appropriate start fill given a set of operating conditions.

The two experiments have also emphasised that the evaporation rate is of significance and should not be considered constant over the year. For this reason also, the start fill is not fixed at an optimum, but will be dependent on the outside air conditions at the time the experiment is operated. This shows that it is relevant to consider batch planning as an independent phase, as it is not a single solution, but should be repeated not only based on the strain of interest and the operating conditions but also external factors such as the outside air temperature and relative humidity.

Despite the success of the batch planning, there are some limitations to the model prediction in this strain. The complex fluid properties are not captured using the simple viscosity model. This has been thoroughly investigated by looking at both on-line and off-line viscosity trends. It is very difficult to capture such complex phenomena. Despite the issues predicting the viscosity in the second half of the batch, the start fill prediction was good enough to capture the process trends and identify a suitable start fill. It is a concern however, for future use of the model for this strain, and it should be considered in control strategy development where the model is applied. For example, it is seen already in this planning work that the modelled feed rate at the end of the batch time is too high. Future work includes a detailed investigation of the on-line viscosity, off-line viscosity and additional studies of the morphology in order to better understand the complex rheological properties exhibited by this strain. In general, focus should also be put on an understanding of the relationship between the on-line and off-line viscosity measurements in the system.

Finally, it has been observed that in tanks 3 and 4 it is not possible to maintain the dissolved oxygen concentration set point in the final hours of the batch. This is due to high viscosity and low oxygen transfer in the system. This will be a consideration for the model-based controller as it is not expected that the model-based controller will prevent oxygen limitation if the DO controlled operation does not.

9 Model-based control strategy

9.1 Introduction

A novel model-based control strategy is developed in order to drive a process to maximum tank fill, subject to the oxygen transfer limitations in a given system. A summary of the control strategy development process and the objectives is provided in detail in Chapter 7. The oxygen transfer rates are dependent on the strain of interest and the viscosity of the fermentation broth, as well as the process conditions. In this work, four sets of process operating conditions are tested in order to challenge the control strategy. In order to achieve the desired tank fill in the defined process time, it must be ensured that the start fill is appropriate for the given operating conditions. This has been discussed thoroughly in Chapter 8.

The model-based control strategy is a form of predictive control, whereby the model is used to simulate the system to the end of the batch time at every time point. Based on the current mass, and the predicted final mass, the feed rate is manipulated to drive the process to a target fill at the target end time. In addition to the feed rate, the head space pressure is manipulated using a simple rule-based approach. The control strategy is applied on-line in two experiments at 550L scale.

A mechanistic model-based control strategy is desirable, as it also allows for future development, and extrapolation to other scales, strains and processes by adaptation of the model. As described in the literature review, it is also a benefit that this method allows for a user to obtain insight into the process by application of the method. It is possible to see how the current process operation compares to the expected process development by the model simulation for example. It is also possible to continually develop the model as additional data becomes available, and new process knowledge is obtained. For industrial application, it is highly valuable to have a mechanistic model available for such process development work.

9.2 Methods

The following section provides a fermentation process description, model description, control algorithm description, and the analytical methods applied.

9.2.1 Process operation

The process operation is defined in Section 8.2.2.

9.2.2 Process conditions

The process conditions are defined in Section 8.2.3.

9.2.3 Off-line sample analysis

The procedures for sample analysis are defined in Section 8.2.6.

9.2.4 Controller algorithm

The concept of the controller is to simulate the process model until the end of the batch time in order to identify if the current batch fill is expected to be above or below the fill target. This simulation is achieved by solving the feed rate necessary to achieve the dissolved oxygen reference trajectory, as shown in Figure 9.1. Adjustments are then made to the feed rate to drive the process towards a desired tank fill (410 kg) in a set process time. This is by adjusting the next feed rate predicted by the model by the error in the final fill divided by the time remaining of the batch. In this way the controller is a form of predictive control, as the model simulates the process to the end of the batch, and then the next control step is implemented. This process is iterated every hour, and is referred to as the supervisory layer.

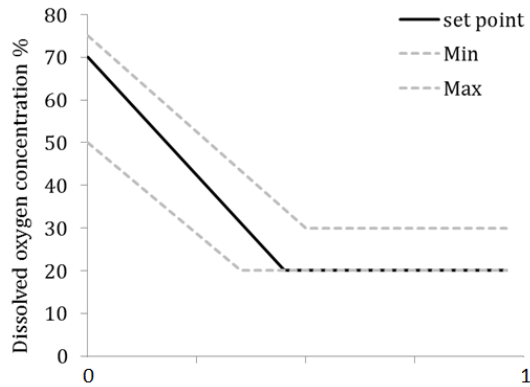


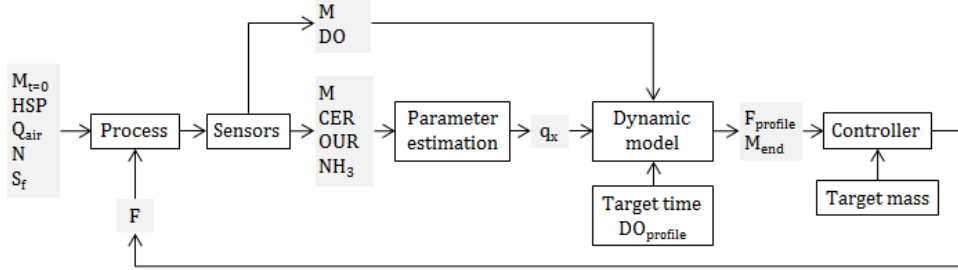
Figure 9.1: Dissolved oxygen control set point ramp, and the maximum and minimum boundary which are required for the regulatory layer.

In order to solve the model every hour, this requires that the initial conditions are specified. The mass and dissolved oxygen concentration is measured directly from the process, and in this way the true initial condition is known. The biomass concentration is not measured on-line, and therefore the mechanistic model based monitoring tool is applied, as described in Chapter 6. The biomass concentration is very central to the prediction of the other model parameters, as it affects the viscosity and therefore the oxygen transfer rate, so it is important for the control strategy that this is the best possible estimate. The other states, for substrate concentration and product concentration may also be found in this way, however they have no impact on the model prediction of the oxygen transfer or mass trajectory in future, and so they are not used in the controller implementation.

When the regulatory layer is also added, the controller will respond to adjust the feed rate applied to the system if the dissolved oxygen goes outside of a specified dead-band. The calculated feed rate from the supervisory layer is adjusted by a proportional controller, in order to avoid oxygen limitation. This regulatory layer is only applied in a specified dead band, between a maximum and minimum dissolved oxygen concentration, as shown in Figure 9.1. This is because it is not necessary to strictly regulate the oxygen profile, but just to avoid over and

under feeding. This may occur due to process model mismatch, but should be avoided in order to maintain a productive process.

Model-based control



Model-based control with regulatory layer

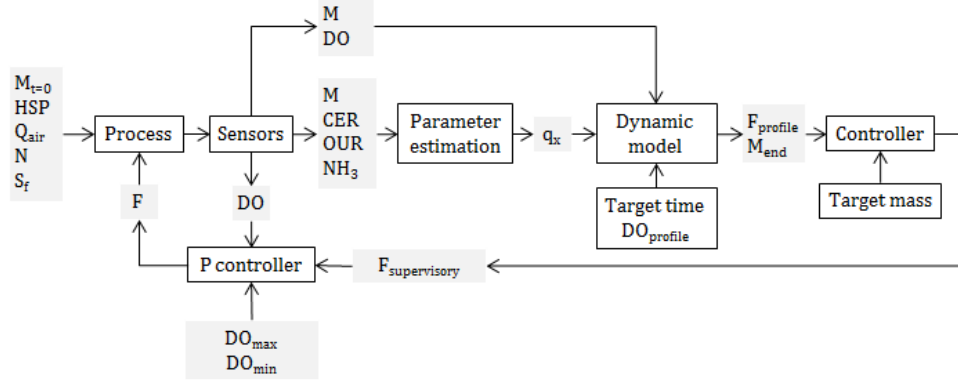


Figure 9.2: Diagram showing the control structure for the model based controller (top) and the model-based controller with regulatory layer (bottom).

In addition to the feed rate, the control strategy includes rule-based control of the headspace pressure set point. This is due to observations that the rapid headspace pressure increase at the start of the process is unnecessary, as the process is not oxygen limited at this stage. This was discussed in detail in Chapter 8. It can also lead to challenges for the controller implementation, as the dissolved oxygen concentration becomes well over the desired set-point which can lead to a response in the controller to over feed. For these reasons, a rule-based approach is adopted, where the headspace pressure is increased in step increments, as the dissolved oxygen falls below a certain level. It is seen that there is an immediate and predictable response in the dissolved oxygen concentration when a step change is applied in the headspace pressure. This is shown clearly in Figure A6. This makes the headspace pressure a good choice for manipulated variable, and also means that it is easy to implement a rule-based control approach, as the gain is clearly observable from the step tests.

The control strategy is best described in Figure 9.3, where the full controller algorithm is shown in detail. The controller is implemented in MATLAB [The MathWorks Inc., 2013], and consists of timers which run scripts for the measurement data collection, supervisory layer model

simulation, and the regulatory layer. There is a Matlab connection to the OPC server in order to access on-line measurement data, and also to write set points which are then applied to the system via the distributed control system (DCS). In experiment 3, the supervisory layer is implemented and this feed is applied directly to the process. In experiment 4, the supervisory layer and the regulatory layer is implemented, as shown in Figure 9.2.

The controller implementation requires two parameters to be specified; FO_{limit} and K_c . These were defined based on process knowledge, but they provide an area for tuning of the controller response. The FO_{limit} term is applied in order to define a maximum feed rate at the start of the fermentation, based on observed over feeding in the DO controller. It was seen from analysis of past data that the increasing oxygen uptake rate is a reliable measurement to show the exponentially increasing capacity for feed at the start of the fermentation. For this reason the maximum feed rate at the start of the fermentation is defined in terms of the OUR. The parameter K_c is the controller gain, which determines how aggressively the controller responds when the dissolved oxygen concentration is outside of the dead-band. This was also fixed in this work by simple analysis of the response of the reference dissolved oxygen controller response applied in Chapter 8. The sensitivity of the dissolved oxygen concentration to the feed rate at the end of the fermentation was observed. This could also have been found by model analysis.

Start controller:

- ☐ Establish OPC server connection
- ☐ Initialise MATLAB® timer objects
- ☐ Load batch conditions (HSP_{max} , Q_{max} , N_{max})
- ☐ Load model parameters
 - ☐ Pre-allocate size of array for data storage
- ☐ Define $DO_{set}=f(t)$, $DO_{min,set}=f(t)$, $DO_{max,set}=f(t)$, t_{end} and M_{target}
- ☐ **Start timers**
 - ☐ Every 5 minutes: Read data timer
 - ☐ Read and save OUR (mol/h)
 - ☐ Read and save CER (mol/h)
 - ☐ Read and save ammonia flowrate (g/h)
 - ☐ Read and save mass (kg)
 - ☐ Read and save feed flowrate (L/h)
 - ☐ Read and save pH (-)
 - ☐ Read and save DO (%)
 - ☐ Every 1 hour: Supervisory layer
 - ☐ Once feed start criteria are met
 - ☐ Take average of past 1 hour of online rate data (q_o , q_c , q_n)
 - ☐ Update parameter estimation (q_g , q_x , q_p)
 - ☐ Define initial condition for state estimate, X_{t-1} , P_{t-1} , G_{t-1}
 - ☐ Solve state estimate, X_t , P_t , G_t
 - ☐ Take average of ten minutes DO data ($DO_{measured}$)
 - ☐ Define initial conditions for model prediction ($DO_{measured}$, $M_{measured}$, X_t)
 - ☐ Solve dynamic model from $t \rightarrow t_{end}$ st. $DO_{profile}$
 - ☐ Determine $F_{t \rightarrow tend}$ and $M_{t \rightarrow tend}$
 - ☐ Error = $M_{tend} - M_{target}$
 - ☐ $F_{supervisory} = F_t - \text{Error}/(t_{end} - t)$
 - ☐ If $t > (t_{end} - 40)$, $F_t \leq F_{t-1}$ (To avoid overfeeding at end)
 - ☐ Implement pressure control (Pressure step change increments)
 - ☐ $DO_{measured} < DO_{set}$ & $HSP_{t-1} < HSP_{max}$, $HSP_t = HSP_{t-1} + 0.05$
 - ☐ Write HSP_t to OPC server
 - ☐ Every 5 minutes: Regulatory layer
 - ☐ Once feed start criteria are met
 - ☐ If $DO \geq DO_{min,t}$ or $DO \leq DO_{max,t}$, $F = F_{supervisory}$
 - ☐ If $DO < DO_{min,t}$, $F = F_{supervisory} + K_c(DO - DO_{set})$
 - ☐ F limits:
 - ☐ $F \leq OUR * FO_{limit}$ (To avoid overfeeding at start)
 - ☐ $F > F_{min}$ (Minimum feed rate)
 - ☐ $F < F_{max}$ (Maximum feed rate)
 - ☐ Feedrate after target time
 - ☐ If $t \geq t_{end}$, $F = F_{t-1}$
 - ☐ If $t \geq t_{end}$ & Error > 5, $F = 0$
 - ☐ Write F_t to OPC server
 - ☐ Manually stop timers at end of batch and save dataset

Figure 9.3: Controller algorithm, described in terms of the timers which are implemented in Matlab R2013b. The values of the controller parameters FO_{limit} and K_c are not provided for confidentiality reasons. They were determined based on experience with the process.

9.3 Results and discussion

In Chapter 8 the start fill calculation has been verified by experiments, and we also have a reference operation performance for the controller. This can then provide a reference case with which to compare the results of the model-based control strategy. Two experiments were conducted using the model-based control strategy as described in Section 9.2. In the following results section, the experiments are described by an experiment number as follows:

Experiment 1: DO controller trial (reference)

Experiment 2: DO controller trial (reference)

Experiment 3: Model based control

Experiment 4: Model based control with regulatory layer

9.3.1 Final fill

The results for the final batch fill are shown in Figure 9.4. Experiment 3, tank 1 is not shown, as this batch was lost due to technical issues, not related to the controller implementation. It is seen that the variation in final fill is reduced to within 5% of the target, and only under filling occurs. In the DO controlled batches both over and under filling of around 10% of the target is seen, and as previously described, the level of over filling would have been greater if there was not manual intervention. This shows that the combination of the calculated start fill and the model-based control is successful to reliably target a desired tank fill. The variation in the final fill is reduced by 74% compared to the dissolved oxygen controlled batches.

There is a difference in the results for the model based control (experiment 3) compared to the model based control with regulatory layer (Experiment 4). The target fill is achieved in the designated batch time with an accuracy of 0.5% for all tanks in Experiment 3, whereas there is under filling of around 5% in Experiment 4. This is due to the trade-off between the fill and the oxygen limitations.

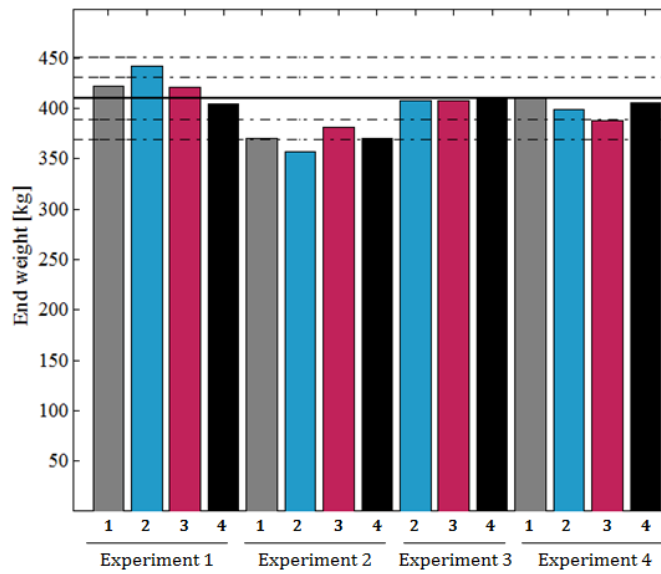


Figure 9.4: End weight of the four experiments conducted. The target fill is shown as a solid line, and 5% and 10% deviations from the target are shown as dashed lines.

9.3.2 Dissolved oxygen concentration

Figure 9.5 shows the dissolved oxygen profiles for experiment 3 and experiment 4, where Tank 2 is shown for comparison. The two different controller implementations, with and without the regulatory layer, may then be directly compared. Beneath the dissolved oxygen profile, the corresponding feed rate profiles are shown.

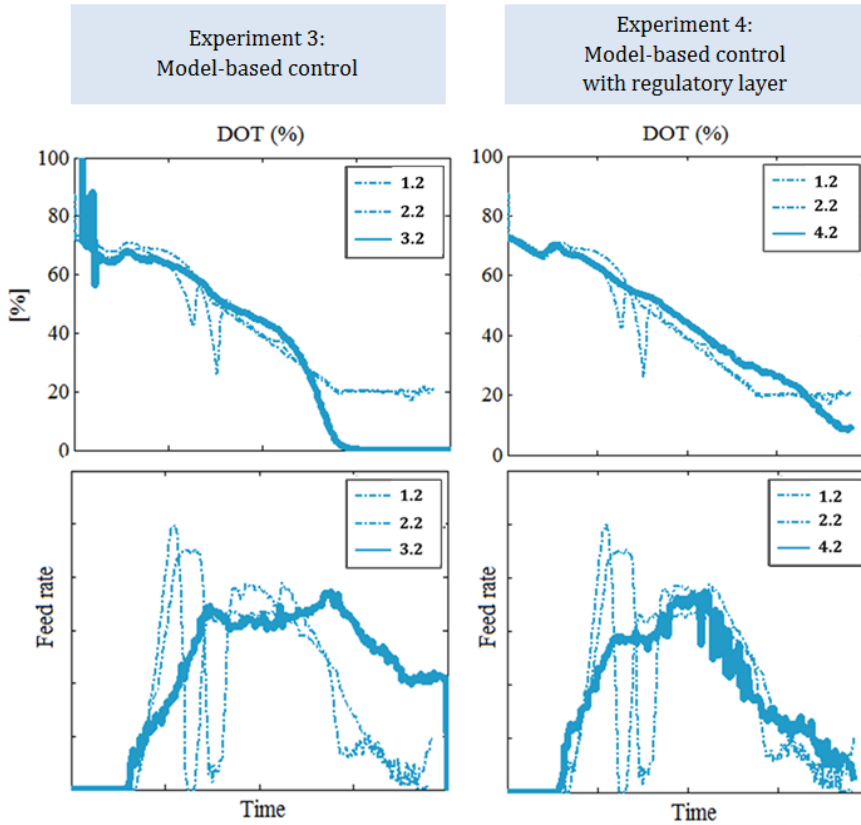


Figure 9.5: Dissolved oxygen profile(top) and feed rate (bottom) for model based control strategy (experiment 3, left) and the model based control strategy with regulatory layer (experiment 4, right). The results show that the addition of the regulatory layer successfully adjusts the model-calculated feed rate in order to avoid oxygen limited conditions.

In experiment 3, the feed rate is solved by the model and applied directly to the system. The feed rate is very accurately determined for the first three quarters of the batch time, with the trend following around the dissolved oxygen set point. However, it is seen that in the final quarter of the batch time, the feed rate is too high for the oxygen transfer rates in the system, resulting in dissolved oxygen limitations. The calculated feed rate is significantly higher than that applied when using the dissolved oxygen controller. As previously explained, the aim is not to achieve a certain oxygen profile, but to avoid complete oxygen limitation, and therefore the feed rate should be close to that of the DO controller. When the regulatory layer is incorporated into the controller algorithm, the feed rate is successfully reduced at the end of the batch in

order to avoid limitation conditions.

It is possible to adjust the DO_{max} and DO_{min} ramps in order to adjust at which point the regulatory layer will be applied. It is also possible to tune the controller gain, in order to respond faster and more aggressively to a value outside of the DO_{max} and DO_{min} ramps. The response of the controller in Experiment 4 is considered to meet the purpose of this exercise, as the mass target is met to within 5%, and complete oxygen limitation is avoided.

The response of the controller can however be manipulated by the user depending on the requirements. For example, it could be considered to increase the controller gain, as it is seen that the dissolved oxygen concentration continues to fall below 20% saturation. With a larger gain, this could be avoided. Although, as mentioned, this would also lead to larger errors in the final mass, and it is not clear that the product concentration is significantly affected by operating at 10% dissolved oxygen concentration at the end of the batch compared to 20%. This is an area of future development for this method, to better understand this trade off, and define a more analytical solution to the balance between product concentration and total mass.

An additional benefit of the model-based feed rate control is that the initial over feeding is avoided. It is seen that the DO controller feeds aggressively in the initial phases of the experiment, since the oxygen level is higher than the set point, however there is limited biomass at this stage, and therefore increasing the feed rate is not the correct action at this stage. There is then an overshoot in the dissolved oxygen concentration when the biomass concentration has increased, and the substrate in the system is consumed. This phenomenon is not experienced with the model-based controller applied in either experiment 3 or experiment 4. This is due to the incorporation of the FO_{limit} term, which was added in order to control the feed rate in the initial batch hours.

As described previously, the concept of the controller algorithm was to reduce the variance in the mass, and as a trade off this would lead to an increase in the variance in another variable. This was defined to be the dissolved oxygen concentration. However it is seen that in fact the variance in the dissolved oxygen concentration is acceptable, and the trend is more similar between the batches for the model based controller than the dissolved oxygen controller. This is due to the overfeeding experienced in the initial phase of the fed-batch process, as previously described. Overall, the results of this control algorithm are seen to be highly favourable; achieving a low variance in the final mass, achieving the target fill to within 5% under the target, and resulting in acceptable variance in the dissolved oxygen profile.

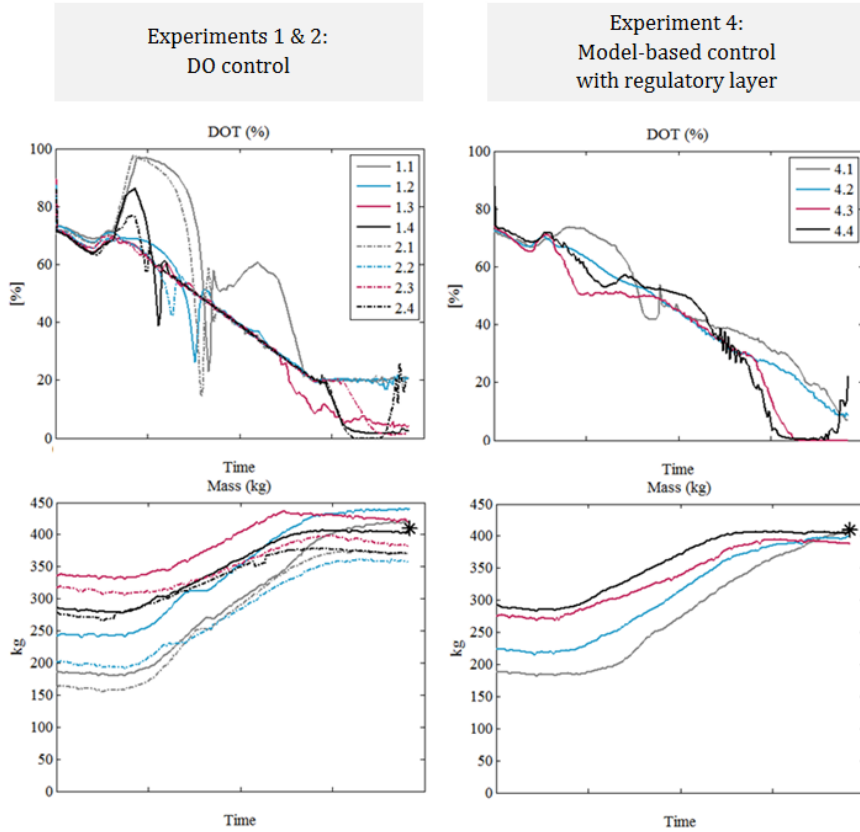


Figure 9.6: Dissolved oxygen profile(top) and mass (bottom) for DO control experiments (Experiments 1 and 2, left) and the model based control strategy with regulatory layer (Experiment 4, right). The results show that the model based controller with regulatory layer successfully targets the target fill, shown with an asterisk (*), and also provides a desirable dissolved oxygen profile.

9.3.3 Headspace pressure

In addition to the feed rate, the headspace pressure set point is also defined by the model based controller. This is due to the observation of very high dissolved oxygen concentrations in experiments 1 and 2 in the beginning of the batch. This occurs for tanks 1 and 4 when the headspace pressure set point is reached using a defined ramp after the feed is started. This high dissolved oxygen concentration is not necessary, since at this stage of the fermentation the process is limited by the biomass concentration. A high dissolved oxygen concentration at this stage is considered a waste of energy, as the increase in pressure also results in an increase in the compressor power required. As is seen in the methods section, the headspace pressure is increased in increments every hour if the measured dissolved oxygen concentration is below set point. In this way the pressure is used in order to increase oxygen supply only when it is required. The result of this is seen in Figure 9.7. The dissolved oxygen control experiments are seen on the left where the pressure is seen to increase in a fixed ramp for tanks 1 and 2. On the right, the headspace pressure is seen to increase only as needed. This results in a very late rise of pressure to the maximum value in tank 1, as the measured dissolved oxygen is within the bounds. The fact that the headspace pressure rises this late may suggest that the start fill was too high in this case, and there was capacity for more feed addition to the system. However, the response of the controller is suitable in this case, as the pressure rises as a response to oxygen becoming limited, but not before it is necessary.

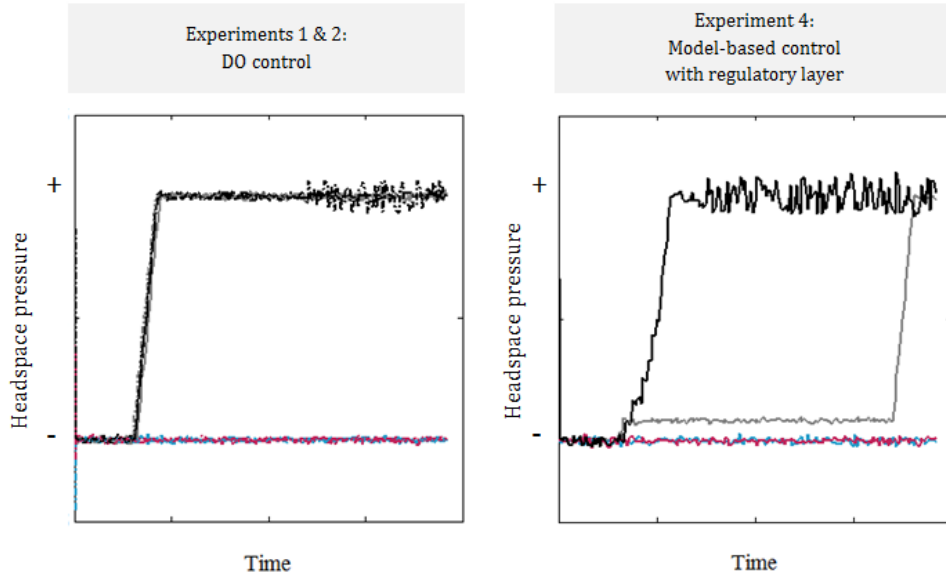


Figure 9.7: Headspace pressure measured for the DO controller experiments (left) and for the model based control with regulatory layer (right). The model based controller also includes adaption of the headspace pressure setpoint. This is to avoid excessively high dissolved oxygen concentrations in the beginning of the batch.

9.3.4 Product concentration

The overall objective for the control strategy is to maximise the total product mass from the system, so it is important to ensure that applying the model-based control strategy does not have a detrimental effect on the product concentration. Figure 9.8 shows the final measured product concentrations for all experiments and all tanks in this work, arranged by tank number for easier comparison. The results show that there is no significant effect, either positive or negative, on the product concentration between the control strategies applied in this work. There is of course batch-to-batch variation in the concentration achieved, but there is no trend which suggests the application of the model-based control strategy has a negative affect on the product concentration which can be achieved for a given set of process conditions. The single outlier is seen in Experiment 2 batch 2, where a high product concentration is measured. This can not be specifically explained. Overall the standard deviation as a percentage of the mean of the product concentration achieved for tanks 1-4 are 4.1%, 7.8%, 1.0% and 5.0% respectively. There is therefore a low variance in the concentrations achieved when applying the different control strategies.

The low product concentration variance means that it would be expected that with repeated application of the control strategy, the final total product achieved per batch would have a lower variance, due to the lower variance in the final fill. The benefit of reducing variance from a control perspective is that it is then possible to be closer to the optimal [Lübbert and Jørgensen, 2001]. In this case it is hoped that the reduced variance in the final mass will allow a more consistent and optimal final total product to be achieved.

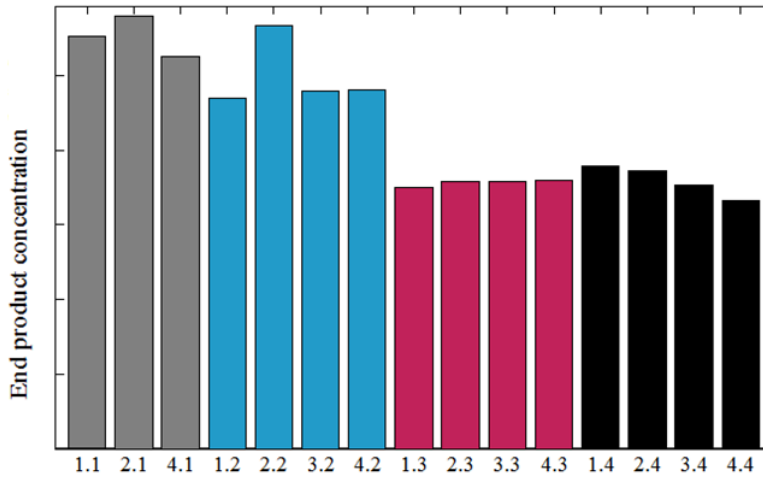


Figure 9.8: The end product concentration for experiments 1-4 and tanks 1-4 where the results are arranged by the tank number for easier comparison. The standard deviation on the mean for the final product concentrations is 4.1%, 7.8%, 1.0% and 5.0% for tanks 1, 2, 3 and 4 respectively. This shows that the product concentration is not significantly affected either positively or negatively by the model based control strategy.

9.4 Conclusions and future work

In this work a novel control strategy was developed in order to maximise the fill in aerobic fed-batch fermentation processes. The target fill was achieved accurately, and with high reproducibility when tested with different process operating conditions, and different meteorological conditions. This is achieved by targeting a desired tank fill in a specified time, whilst avoiding oxygen transfer limitations during progression of batch operation. It also included a model-based batch planning stage where the initial batch fill is defined. This is a novel control concept.

The control strategy has been rigorously and successfully tested in pilot scale (550 L) studies by performing in total 16 runs, including the reference control strategy, under four different industrially relevant process operating conditions. The results for the model-based control strategy with regulatory layer has been shown to meet the objective, with the target mass reached to within 5% under the target, and over filling was avoided. The reduction in variance of the final fill of over 74% is not only beneficial for the final product mass, but also for planning and scheduling in a multi-product facility, where downstream operations must also be scheduled. With a reliable product mass flow to recovery, resource allocation is more predictable.

In conclusion, focusing on final batch fill is considered to be a highly valuable control objective, especially in industry where highly optimised strains are applied. In the fermentation control literature, there are often references to control methods to obtain high product concentration, however in highly engineered industrial hosts we feel there are greater benefits in optimising total mass, than optimising product mass fraction. In this way it is not only possible to achieve maximum product mass in a batch operation, but it allows for predictable product mass to the downstream operations, which allows for improved scheduling and capacity planning.

Future work for this project includes a consideration of gas hold up in the fermentation system, to determine the optimal maximum fill, which in this work has been fixed based on prior experience. This adds a lot of complexity to the problem, as the gas hold up is hard to define, and also changes over time based on the processing conditions and the rheological properties of the fluid. It is however very important to consider this as an approach to determine the optimal target fill in a dynamic process. In this work, the target fill was fixed, and as discussed, this also has the added benefit for the full process, that the downstream scheduling operations are more simple when the total mass is fixed. However, it could lead to an even better use of available capacity if the maximum fill could be solved for a system on-line.

Other developments to this method could be implemented if there is a better understanding of how the metabolic rates are affected by the processing conditions. This specifically relates to the relationship between the dissolved oxygen concentration and the rate of product formation. There is an opinion that the level of dissolved oxygen should be kept above zero in order to avoid oxygen limited conditions, as this will affect the product formation. There is however no specific understanding of this which may be implemented in the model. Also this is hard to verify, for example Figure 9.5 shows that in Experiment 3, oxygen limitation occurs, however in Figure 9.8 there is no reduction in the product concentration. This lack of mechanistic understanding limits the application of the model-based approach. For example, it is not possible to make more defined rules about the trade off between the final fill and the product concentration.

If this knowledge was available, it would be possible to run a full process optimisation, where the objective is explicitly to maximise the total product achieved at the end of the batch. The optimal dissolved oxygen profile would then be obtained, and the trade-off would be implicit in the optimisation problem. If we implement this full process optimisation using the mechanistic

model in this work, we have to specify a minimum dissolved oxygen level, and then the solution is to feed at the maximum rate until the maximum tank fill is reached, and then the feed is set to a minimal level in balance with the evaporation rate until the end of the batch. In conclusion, if the model complexity is increased, it can also allow for more optimal control algorithms to be developed.

Overall conclusions and suggestions for future work

For each chapter in this thesis, the conclusions and relevant future work are described. This chapter brings together these considerations, and also provides a broader perspective on the different approaches which could be considered to improve on this work in the future.

Modelling and monitoring

With respect to modelling of fermentation processes, two approaches have been considered; multivariate analysis and mechanistic modelling. Multivariate analysis provided a faster approach to modelling of the production scale process, for which there was no available process model. The resulting model provided possible leads for future process optimisation, by analysis of the regression coefficients. There is a huge potential for these data analysis tools to be applied in industry, where there is a vast amount of historical data available. There is however, a genuine challenge in organising, aggregating, and aligning the relevant data, due to the multiple systems used for data collection and storage. If the historical datasets were more readily accessible, such tools may be applied more routinely in industry. The multivariate analysis methods are well established, and the focus going forward should be on the integration of data from different sources and the development of front end tools for users. In this way the tools available, and for example the methodology developed in Chapter 4, may be applied more widely.

Despite the benefits of the data-driven approach for data mining and providing process insight, a mechanistic model is desirable for control purposes, as it is predictive and better able to be applied to conditions outside of those used to develop the model. In this way, a mechanistic model approach is considered more flexible and robust for many applications, as shown graphically in Figure 5.1. In order to re-use and further develop existing mechanistic models it is important to employ good modelling practices and document the model thoroughly in order to allow for sharing of model code [Sin et al., 2009]. A part of good modelling practice is also to conduct uncertainty analysis in order to determine the limitations of the model, and to interpret the model results in an objective way. This also allows the user to judge for what applications the model may be applied.

In this work it was shown that there was some lack of understanding which led to a poor prediction of some model outputs. Firstly, the uncertainty analysis showed a high uncertainty in the model parameters related to yield and maintenance terms. Considering the benefits and disadvantages of both black-box models and mechanistic models, hybrid modelling may provide a very promising approach to modelling of biological systems [von Stosch et al., 2014]. In this way, the phenomena which are well understood, may be described by a fundamental mechanistic model, and the model parts for which there is a lack of understanding may be modelled using a data driven approach. This results in a model which has better extrapolation capabilities than a black-box model, due to the fundamental model basis. It may also allow for better prediction capabilities, by replacing an area of the model which is poorly described with a data-driven model. This approach has been successfully applied to fermentation processes, and a detailed

description of the field is described in von Stosch et al. (2014).

In this work, it could be interesting to compare the results of the monitoring tool described in Chapter 6 with a hybrid modelling approach, where the rates are defined by a purely data-driven model. In an industrial setting, this may be particularly valuable, where there may be multiple production strains being tested in the same equipment. In this way, a hybrid model may be well suited, as the physical model is defined by the reactor configuration and the physical process parameters and may therefore be described mechanistically. In contrast, the strain-specific parameters are unknown for each strain, and these may then be identified using past experimental data. A hybrid approach may be well suited for expanding the model application range to new strains in this way.

When discussing the broader application of mechanistic models, the scale of application must be considered. In this work, the mechanistic model was applied at pilot scale only, however future work should consider application at production scale, to allow for application of the batch planning tool (Chapter 8) and the control strategy (Chapter 9) to the production scale processes. A challenge related to modelling at larger scale is the lack of understanding of the heterogeneities which may be influencing the system [Gernaey et al., 2010]. This type of understanding may come through specific experimentation, ideally in collaboration with computational fluid dynamics (CFD) studies. This is outside the scope of this work, however in future as this understanding becomes available, it may guide model development, to allow a better model prediction at larger scale. For example, CFD studies may aid the development of appropriate compartment models to capture the gradients in these large systems using mechanistic models.

Control

The introduction to this project included a literature review covering some approaches which have been employed in order to control the feed rate to a fed-batch fermentation process. Over the development of this work however, an approach has been developed which does not strictly fit to any of the control strategy types described in the review. Although the method is a model-based approach which uses a future prediction in order to solve the next control action, this approach is not model predictive control, as we do not solve an optimisation problem at each iteration, and there is no penalty on the magnitude of the control action taken. The control strategy was developed as a natural continuation of the monitoring strategy, where the model was already implemented on-line in the pilot facilities at Novozymes A/S, and the controller algorithm then utilised this on-line application, and extended the simulation time until the end of the batch time. This has shown to be a successful method, and has met the objectives for this work, however control engineers will likely not be completely satisfied with the very objective and applied control strategy development process.

As discussed in Chapter 9, there are limitations to the model which limit the ability to implement full process optimisation on-line. This is mainly due to a lack of understanding of how the dissolved oxygen concentration and the substrate concentration affects the metabolic rates in the system. If the model is applied for process optimisation in the current form, the feed will be added at the maximum rate until the oxygen concentration goes to zero. This is since there is no relation between the oxygen concentration, the substrate concentration and the rates. This requires a more complex model, however there is a lack of understanding in order to implement this. Ideally, with a more complex model approach, the aim would be to implement a full process optimisation where the feed rate is solved at each iteration in order to maximise the total product achieved in a fixed time.

Another improvement to the control strategy is to account for gas-hold up in the optimisation. As described in Chapter 9, it would be beneficial to have the maximum mass in the system determined on-line based on the gas hold up, as the maximum fill in a fed-batch system is dependent also on the gas-hold up in the system. This adds significant complexity to the control problem. Gas hold up is challenging to define, and there is no available measure for this parameter. There is a current lack of understanding of the factors affecting the level of gas hold up, so modelling is also very challenging.

Finally, this work has also discussed the importance of accounting for evaporation in an industrial fed-batch process. With relation to this, if a model predictive control approach was implemented, all operating conditions could be optimised on-line, including the headspace pressure and the aeration rate, which could also allow the control problem to be extended to account for evaporation. With the ultimate goal to achieve the maximum total product mass in a fixed process time, this could also then allow for on-line adaptation to manipulate the evaporation rate. For example, for a process which is almost at the maximum tank fill with remaining process time, the head space pressure could be reduced to increase evaporation and allow for additional feed to be added, and converted to product. This would of course lead to a decrease in the oxygen availability, but with a comprehensive process model this trade-off could be addressed.

Appendices

Supplementary material: Chapter 3

Table A1: Adaptive control literature review. Asterisk indicates a variable which is calculated (soft sensor approach).

Control Basis	CV	MV	MeV	Objective	D	Testing
(Oliveira et al., 2004) Model Reference Adaptive Control (MRAC). Indirect adaption refers to the estimation of unknown process parameter (θ). The reference model is then tracked by tuning of the controller parameters. 30% DO set point.						
$F = OTR - \left(\frac{DO_{SP} - DO}{\tau} \right) \cdot \frac{V}{\theta + DO}$	DO	F	DO, OTR	Max OTR	$k_L a$ step	Pilot simulation
<i>Where θ is solved for as an unknown and τ is tuned.</i>						
(Jenzsch et al., 2006) X_{Total}^* trajectory calculated from an optimal μ trajectory. ANN trained on 26 batches used to estimate X from 3 inputs. Control law parameter α then adapted based on the error between the X_{Total}^* and X_{Total} set point.						
$F = \frac{\mu_{set} \cdot X_{Total}^*}{(Y_{XS} - \alpha) \cdot S_f}$ <i>Where α is tuned based on X error</i>	X_{Total}^*		CER, OUR, Base	Control X_{total} trajectory	F	15L
(Duan et al., 2006) ANN network trained on pH and DO data used to define 2 physiological states; substrate starvation or excess. Feed rate adapted based on this state of the system. Final biomass obtained by ANNPR controls increased about 47–55%, and the cultivation time was also shortened 20–43%.						
$F(k) = F(k-1)(1 + \delta T)$ <i>Where δ denotes positive step size T defines increase/decrease in pump setting</i>	ANN state	F	DO, pH	Max X, Min Xal	-	5L
(Soons et al., 2006) MRAC method based on a dual substrate model compared to a reference μ model for <i>B. pertussis</i> cultivation. Controller gains, K_1 and K_2 are updated every minute based on the estimated states (μ , X and V).						
$F_{1+2} = \frac{ac + bd}{a F_1 + b F_2} X^* V + K_1 (\mu_{set} - \mu) + \int_0^t K_2 (\mu_{set} - \mu) dt$ Two substrates, given here as F_{1+2} . Where a, b, c, and d are constants, calculated based on the model equations and the desired growth rate set point. K_1 and K_2 are tuning parameters.	μ	F	DO, OUR* X^* V	Constant μ	DO, F	5L
(Jenzsch et al., 2007) Feedforward ANN with single hidden layer used to determine the total cumulative CER set point based on X, induction time, and μ . A form of adaptive PI control then updates feed rate. tcCER control variable chosen as it is measured in line, no time delay, no heterogeneity issues with measurement and robust to X deviations/ μ deviations.						
$F = \gamma F_{reference}, \text{ Where } \gamma = 1 + k_1 \Delta tcCER + k_2 \int_{ts}^t \Delta tcCER dt$ Where tcCER is total cumulative CER. k_1 and k_2 tuning parameters.	tcCER	F	CER	Reproducible batches	-	15L
(Oliveira et al., 2005) Paper describes both a MRAC and an integral feedback with adaptive control method, which is considered more effective and is described in the table. Control was found to be more stable with an OTR estimate rather than measurement due to delays.						
$F(t+1) = F(t) - \frac{T}{\theta_t \tau_c^2} (DO_{set} - DO)$ Where T is the time step, θ_t is the unknown state variable, and τ_c^2 is the controller time constant.	DO	F	DO	Max OTR	DO, OTR noise	50L

Table A2: Model predictive control (MPC) literature review. Asterisk indicates a variable which is calculated (soft sensor approach).

Optimisation problem	Objective	CV	MV	MeV	Testing
(Zhang and Lennox, 2004)					
Multivariate Partial least squares based on 20 batches used as a soft sensor for biomass concentration. MPLS then integrated into a MPC technique, to control the X trajectory. Tested on a simulated penicillin fed batch process.					
Set point tracking	Reproduce X profile	X*	F	F, Q _{air} Agitator power Substrate temperature, S, DO, V, T, pH, heat generated	simulation
(Kuprijanov et al., 2013)					
Adaptive MPC used to follow optimal X profile from a 'golden batch'. Offline X and S measurements were required as inputs to the system. Aim of the paper is to show the MPC method implemented simply in a standard industrial control system. <i>E. coli</i> system.					
Set point tracking	Reproduce X profile	X	F	M, X, S	10L
(Craven et al., 2014)					
Non linear MPC, applied to mammalian fed batch process. In situ ramen spectrometer used for metabolite measurement. Fixed glucose concentration set point of 11mM with a planned step change to 15mM after 100 hours. Simulation studies to tune the controller before lab testing at 15L scale.					
Set point tracking	Maintain S	S	F	S, P	15L
(Santos et al., 2012)					
Non linear MPC. Cost function aims to maximise glucose oxidation, and minimise glucose fermentation at a kinetic level.					
Maximise variable	Max glucose oxidation		F	OTR CER OUR	simulation
(Kovárová-Kovar et al., 2000)					
ANN combined with MPC. ANN used in place of a model for product formation, which is not currently able to be modelled reliably. Objective function to maximise total product and yield of product. Increase yield by 10%.					
Maximise variable	Maximise product yield and total product	P	F	CER F Nitrogen added Glucose utilized Cumulative CO ₂ pH	15L
(Chang et al., 2016)					
Dynamic flux balance model applied to <i>S. cerevisiae</i> . Optimised feed rate and dissolved oxygen set point					
Maximise variable	Maximise ethanol product concentration	P	F DO setpoint	V X P S	simulation

Table A3: Fuzzy control literature review. Asterisk indicates a variable which is calculated (soft sensor approach).

Control Basis or linguistic variables	CV	MV	MeV	Objective	D	Testing
(Zhang et al., 1994)						
Two fuzzy sets were used for different phases of the fermentation. During the initial alcohol production phase a set of rules were used with the error in RQ being the control variable. The next phase, the μ_{\max} was the set point for the system, and the calculated μ compared to the optimal μ_{\max} provided the error input to the fuzzy controller.						
$F = (1 + \alpha)F * (t - 1)$ if $RQ \leq 1$ $F = (1 - \alpha)F * (t - 1)$ if $RQ > 1$ Where α is the output of the fuzzy rules and F^* is the calculated optimal feed rate.	RQ, μ^*	F	RQ, μ^*	RQ=1 (phase1), $\mu^* = \mu_{\max}$ (later phases)	-	Simulation
(Horiuchi and Hiraga, 1999)						
The process is defined as consisting of four phases of operation, where fuzzy logic is applied to identify the current phase. Each state variable, process time, DO, CER and CER total were assigned a trapezoid membership function. F and pH controlled together. Total product increased 6-16% compared to manual control, and has been in operation for two years.						
$F(t) = F_A - a_{FA} + b_{FA} * t$ Where all parameters are fixed for the four different phases of operation and fuzzy logic is used to identify the phase of operation.	-	F, pH_{set}	DO, CER, pH	Reliability and optimal operation (P)	-	Industrial
(Hisbullah and Ramachandran, 2003)						
Three control methods tested with the optimal being a hybrid fuzzy and PI control system with scheduled gain. Reduced oscillations and set point tracking without offset. Set point of specific CER- specific OUR. Requires online biomass measurement or estimate.						
$e = (QCER - QOUR)_{\text{setpoint}} - (QCER - QOUR)_t$ Error then defines the membership rules and feed rate is defined from this.	$QCER - QOUR$	F	X, CER, OUR	Maintain setpoint $(QCER - QOUR)$	$QCER - QOUR$ noise	Simulation

Table A4: Artificial Neural Network (ANN) control literature review. Asterisk indicates a variable which is calculated (soft sensor approach).

Training algorithm	Optimisation algorithm	Inputs	Outputs	Objective	D	Testing
(Chen et al., 2004) Cascade recurrent neural network for fed batch <i>S. cerevisiae</i> process. 2 blocks, where the F and V are used to first estimate DO. The DO is then input to the second block which is used to estimate X. A genetic algorithm is used to optimise a smooth feed profile. Volume can be measured or calculated.						
LMBP	Genetic Algorithm	F, V or V*	X	Max X_{total}	-	Lab
(Peng et al., 2013) Modelled an ANN based on 5 batches, and later added an additional batch to improve performance. Created an optimal trajectory using the Genetic Algorithm. Ran 25 batches using the optimal trajectory.						
Back propagation	Genetic Algorithm	Time, pH, DO, T, turbidity	P	Max P	-	5L
(Ferreira et al., 2001) Multilayer ANN used to interpret biosensor signal for glucose and sucrose. Control of an alcohol fed batch fermentation using ANN output and PI controller to control substrate concentration.						
Conjugate gradient	-	F, S (biosensor output mV)	S	Meet S set point	-	Simulation

Table A5: Probing control literature review.

Control Basis	CV	MV	MeV	Objective	D	Testing
(Johnsson et al., 2013)						
Power spectral density of the DO signal used to determine the frequency content, which is used directly in the control algorithm. Aim to maximise the oxidative metabolism, whilst avoiding overflow metabolism. Parallel system controls DO using N and Q_{air} but with slow response, so it does not interact with feed controller.						
$F(t) = \frac{K}{h} (C_{sp} - C_k) + \frac{h}{T_i} \sum_{1 \leq j \leq k} (C_{sp} - C_j)$ <p>Where C is the frequency content of the DO spectral data, calculated from the PSD#.</p> <p>Gain scheduling:</p> $K = K_0 \sqrt{\frac{F_{k-1}}{F_{min}}}$	DO	F, Q_{air} , N	DO	Max and maintain oxidative Qs	F, $X_{(t=0)}$	550L
(Henes and Sonnleitner, 2007)						
Systematic underfeeding. Relative change in DO used, not absolute values, for robustness. Standard exponential feed equation used as starting point, and adaptive strategy used to manipulate μ (± 1 -5%) in the equation. Applied to <i>S. cerevisiae</i> , <i>E. coli</i> and <i>P. pastoris</i> . DO signal can be sensitive to antifoam addition.						
$F(t) = \frac{\mu \cdot X_0 \cdot M_0}{Y_{XS} \cdot (S_f - S_0)} e^{\mu t}$ <p>Where μ is adapted based on underfeeding response:</p> $F(t) = F_0 e^{\mu t}$	-	F	DO	Max and maintain oxidative Qs		30L
(Velut et al., 2007)						
Probing control combine with temperature limited operation in order to limit the OUR. Initially probing control alone implemented. When N is approaching the maximum, (hence OTR maximised) temperature control loop introduced. Two SISO loops in cascade. Feed is adapted in proportion to the response, or reduced if no response in DO is measured.						
<p>If $O_{pulse} < O_{reaction}$</p> $F(k+1) = F(k) + F_{dec}$ <p>Else</p> $F(k+1) = F(k) + K(O_{pulse}(k) - O_{des})$ <p>Where</p> $F_{dec} = \text{predefined constant}$ $O_{detection} = \text{detection limit}$ $O_{des} = \text{desired response}$	-	F, T, N	DO, T	Max and maintain oxidative Qs	F, Sf, X	3L

#PSD= Power Spectral Density. Frequency content equations provided in Johnsson et al. 2013.

Table A6: Statistical process control literature review.

Multivariate methods	MeV	Objective	Testing
(Albert and Kinley, 2001) Paper from Eli Lilly and Co. Industrial application of multivariate analysis to provide online process control. Matlab interface implemented for display to the user. PCA applied to 144 batches. 65 high yield batches defined as desirable operation. 5 principle components found to characterise the high yielding batches sufficiently.			
PCA	17 online 53 offline	Follow high yielding batches	Industrial
(Duran-Villalobos et al., 2016) PLS model developed and then applied to future batches using an optimisation function to determine the optimal manipulated variables for the future batch based on their initial conditions. Successful results for two simulation case studies.			
MPLS model	up to 15	Maximise biomass formation/Maximise product formation	Two simulation studies

Supplementary material: Chapter 8

In Chapter 8, the results of the reference DO controlled experiments are discussed. There is a focus on the viscosity which is measured both in-line and off-line. The off-line measurements are modelled using the Bingham model (Equation A1), as it was shown to provide the best fit to the experimental data with two parameters, as shown in Figure A2.

$$\mu = \frac{\tau_o + \mu_p \dot{\gamma}}{\dot{\gamma}} \quad (\text{A1})$$

It is common to apply the Power law to viscosity measurements of fermentation broth, as there is a shear thinning behaviour. This shear thinning behaviour can also be seen in Figure A2. In this work, a wide range of shear rates were sampled for the off-line viscosity measurement, however it is only the lower shear rates which are relevant for the conditions in the fermentation. In future, it would be advisable to use the power law, fitted only to the shear rates in the range relevant for the fermentation system of interest. The power law may be fitted to the data, if only the low shear rate values are included in model fitting. It is shown in Figure A1 that this does not impact the values for the viscosity which are presented in this work, where both methods result in the same apparent viscosity value at the given shear rate.

$$\mu = K \dot{\gamma}^{n-1} \quad (\text{A2})$$

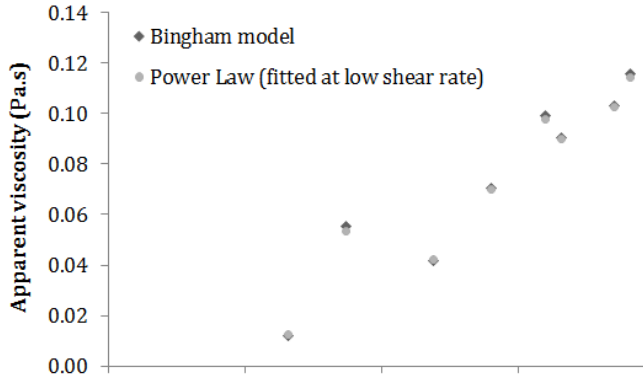


Figure A1: Viscosity obtained by the Bingham plastic viscosity model and the Power law fitted to low shear rate data only. Data shown for Experiment 1, Tank 1.

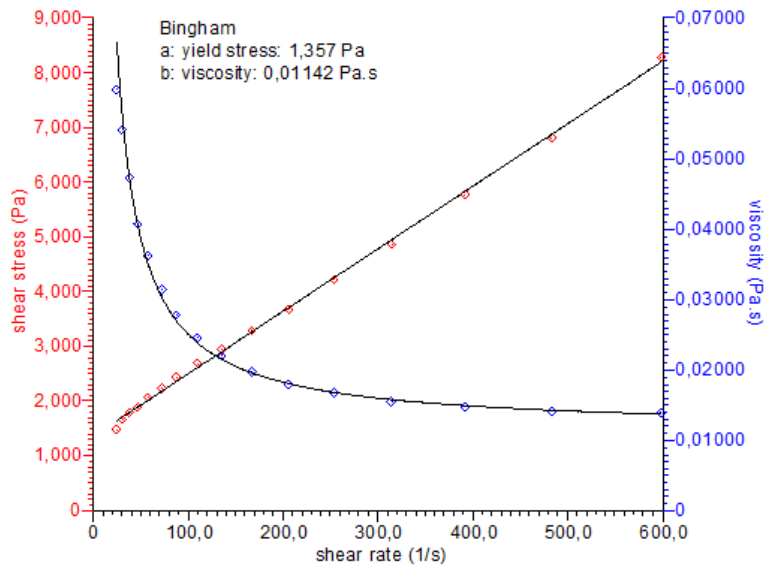


Figure A2: Bingham plastic viscosity model fitted to off-line measured data.

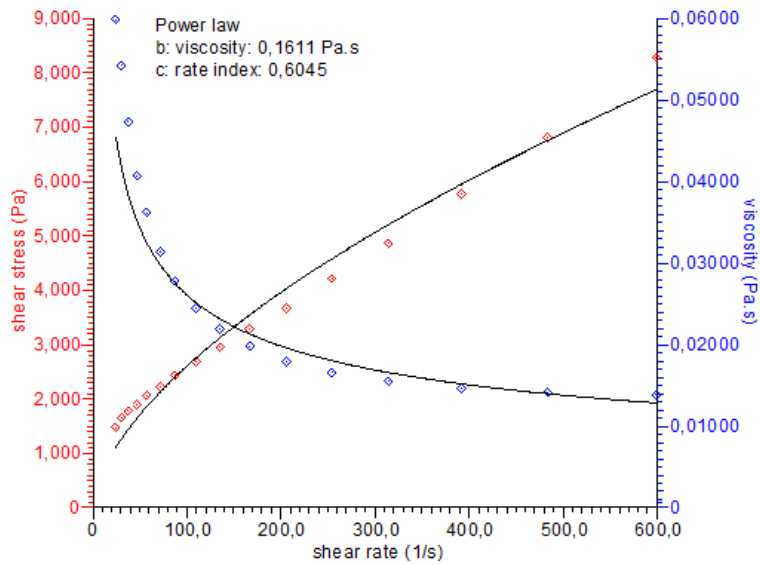


Figure A3: Power law viscosity model fitted to off-line measured data.

When a model is identified which provides a good fit to the experimental data, the model parameters may be investigated to identify trends in the viscosity. This is useful in this work, where two stirrer speeds are used in the experiments, and so there are two different shear rates which represent the conditions in the tanks. For this reason, it is useful to directly compare the viscosity model parameters, as they are independent of the stirrer speed for a given tank. This may provide a more fundamental understanding of the rheological properties of the broth. This may provide some insight into the potential for morphological changes in the system. Figures A4 and A5 show the model parameters for the Bingham model and the Power law model respectively. It is outside the scope of this work to analyse the results in depth, however it is seen in both cases that tanks 1 and 2 have a similar behaviour, and tanks 3 and 4 also. This implies that the stirrer speed has an effect on the rheological property of the broth.

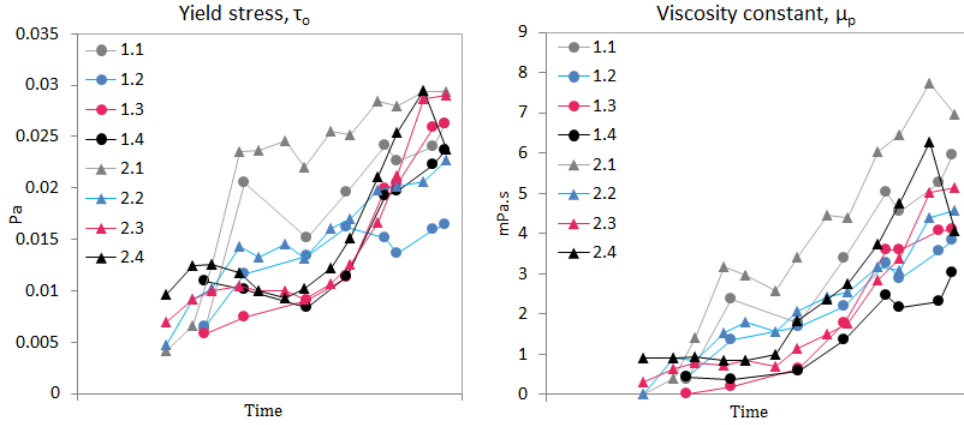


Figure A4: Bingham model parameters for off-line measured viscosity in dissolved oxygen controlled reference experiments.

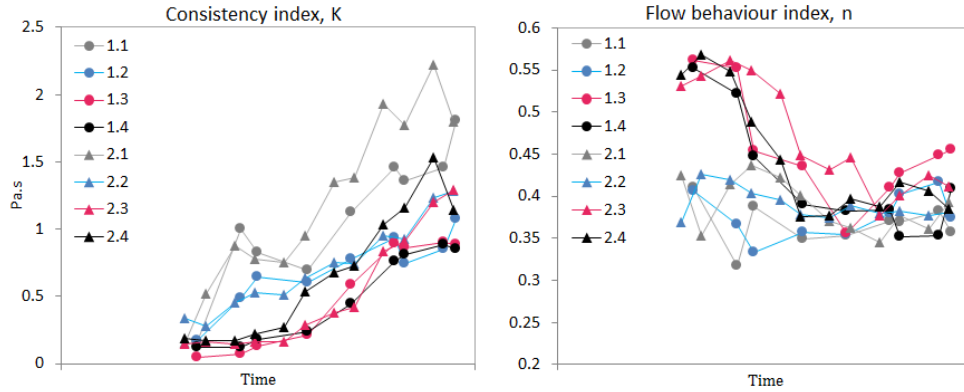


Figure A5: Power law model parameters, fitted for low shear rate data only, for off-line measured viscosity in dissolved oxygen controlled reference experiments.

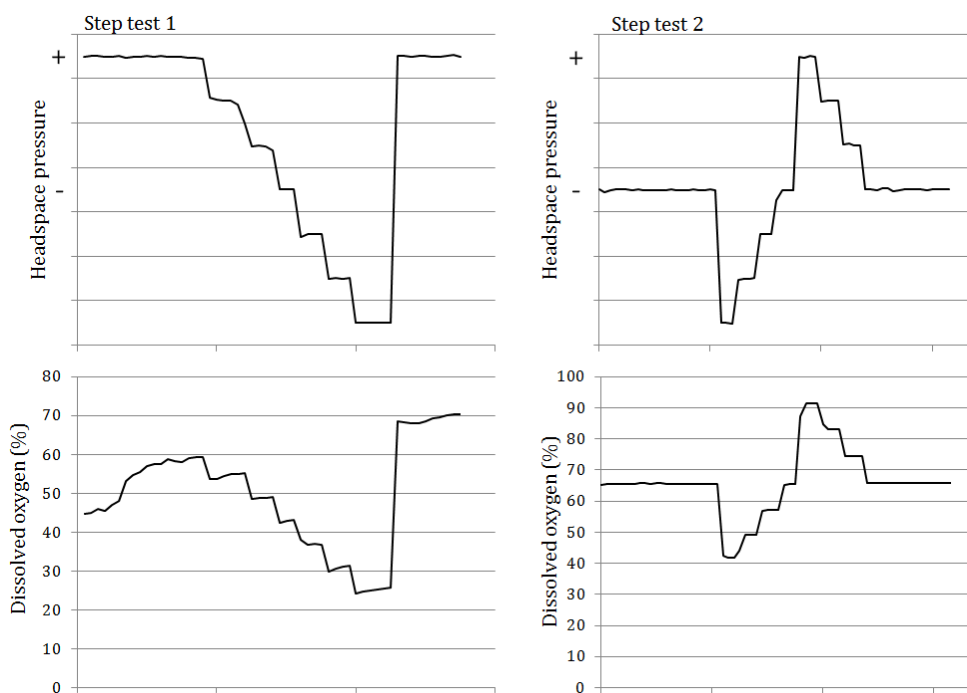
Supplementary material: Chapter 9

Figure A6: Two step test experiments for headspace pressure (top) and response in dissolved oxygen concentration (bottom).

List of publications

Peer reviewed manuscripts

Mears, L., Nørregård, R., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.V., Villez, K. (2015). Multivariate Analysis of Industrial Scale Fermentation Data. *Computer Aided Chemical Engineering*, 37, 1667-1672.

Mears, L., Nørregård, R., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.V., Villez, K. (2016). Functional Unfold Principal Component Regression Methodology for Analysis of Industrial Batch Process Data. *AIChE Journal*, 62, 6, 1986-1994.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.V. (2016). Mechanistic models for Process Development and Optimization of Fed-Batch Fermentation Systems. *Computer Aided Chemical Engineering*, 38, 1311-1316.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.V. (2017). Application of a Mechanistic Model as a Tool for On-line Monitoring of Pilot Scale Filamentous Fungal Fermentation Processes- The Importance of Evaporation Effects. *Biotechnology and Bioengineering*, 114, 3, 589-599.

Mears, L., Stocks, S. M., Sin, G., Gernaey, K.V. (2017). A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. *Journal of Biotechnology*, 245, 34-46.

Mears, L., Stocks, S. M., Albaek, M. O., Cassells, B., Sin, G., Gernaey, K.V. (2017). A novel model-based control strategy for aerobic filamentous fungal fed-batch fermentation processes. *Biotechnology and Bioengineering*, Accepted, doi:10.1002/bit.26274

Conference abstracts

Mears, L., Nørregård, R., Stocks, S. M., Sin, G., Gernaey, K., Villez, K. (2015). Novel strategies for control of fermentation processes. 19th Nordic Process Control Workshop, NTNU, Norway, 15-16 Jan 2015, Accepted for poster presentation.

Mears, L., Nørregård, R., Stocks, S. M., Sin, G., Gernaey, K., Villez, K. (2015). Application of Multivariate Analysis Tools to Industrial Scale Fermentation Data. *Advances in Process Analytics and Control Technology Conference (APACT15)*, Manchester, UK, April 22-24 2015, Accepted for oral presentation.

Mears, L., Nørregård, R., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K., Villez, K. (2015). Multivariate analysis of industrial scale fermentation data. 12th International Symposium on

Process Systems Engineering and 25th European Symposium on Computer Aided Process Engineering (PSE2015/ESCAPE25), Copenhagen, Denmark, May 31 - June 4 2015, 1667-1672. Accepted for oral presentation.

Mears, L., Nørregård, R., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K., Villez, K. (2015). Functional unfold principal component analysis methodology for industrial scale fermentation data. 10th European Congress of Chemical Engineering - 3rd European Congress of Applied Biotechnology - 5th European Process Intensification Conference (ECCE10+ECAB3+EPIC5), Nice, France, September 27 - October 1 2015. Accepted for poster presentation.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.. (2015). Development of an on-line state estimator for fed-batch filamentous fungal fermentations, Recent Advances in Fermentation Technology (RAFT11) Florida, US, November 8-11 2015. Accepted for poster presentation.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.. (2016). On-line state estimation applied to 550L filamentous fungal fermentations, 4th BioProScale Symposium Bioprocess intensification through Process Analytical Technology (PAT) and Quality by Design (QbD) April 6 - 8 2016, TU-Berlin. Accepted for poster presentation.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.. (2016). Mechanistic models for Process Development and Optimization of Fed-Batch Fermentation Systems, 26th European Symposium on Computer Aided Process Engineering (ESCAPE26). June 12-15 2016, Portorož, Slovenia. Accepted for poster presentation.

Mears, L., Stocks, S. M., Albaek, M. O., Sin, G., Gernaey, K.. (2016). Model-based control of a pilot scale fermentation process using a predictive control strategy, 20th Nordic Process Control Workshop. August 25-26 2016, Stockholm, Sweden. Accepted for oral presentation.

Bibliography

- [Åkesson et al., 1999] Åkesson, M., Hagander, P., and Axelsson, J. P. (1999). A probing feeding strategy for *Escherichia coli* cultures. *Biotechnol. Tech.*, 13(8):523–528.
- [Albaek, 2012] Albaek, M. O. (2012). *Evaluation of the efficiency of alternative enzyme production technologies*. PhD thesis.
- [Albaek et al., 2011] Albaek, M. O., Gernaey, K. V., Hansen, M. S., and Stocks, S. M. (2011). Modeling enzyme production with *Aspergillus oryzae* in pilot scale vessels with different agitation, aeration, and agitator types. *Biotechnol. Bioeng.*, 108(8):1828–40.
- [Albaek et al., 2012] Albaek, M. O., Gernaey, K. V., Hansen, M. S., and Stocks, S. M. (2012). Evaluation of the energy efficiency of enzyme fermentation by mechanistic modeling. *Biotechnol. Bioeng.*, 109(4):950–61.
- [Albaek et al., 2008] Albaek, M. O., Gernaey, K. V., and Stocks, S. M. (2008). Gassed and ungassed power draw in a pilot scale 550 litre fermentor retrofitted with up-pumping hydrofoil B2 impellers in media of different viscosity and with very high power draw. *Chem. Eng. Sci.*, 63(24):5813–5820.
- [Albert and Kinley, 2001] Albert, S. and Kinley, R. D. (2001). Multivariate statistical monitoring of batch processes: an industrial case study of fermentation supervision. *Trends Biotechnol.*, 19(2):53–62.
- [Alford, 2006] Alford, J. S. (2006). Bioprocess control: Advances and challenges. *Comput. Chem. Eng.*, 30(10-12):1464–1475.
- [Alves-Rausch et al., 2014] Alves-Rausch, J., Bienert, R., Grimm, C., and Bergmaier, D. (2014). Real time in-line monitoring of large scale *Bacillus* fermentations with near-infrared spectroscopy. *J. Biotechnol.*, 189:120–8.
- [Andersson, 2009] Andersson, M. (2009). A comparison of nine PLS1 algorithms. *J. Chemom.*, 23(10):518–529.
- [Araújo-Bravo et al., 2004] Araújo-Bravo, M. J., Cano-Izquierdo, J. M., Gómez-Sánchez, E., López-Nieto, M. J., Dimitriadis, Y. A., and López-Coronado, J. (2004). Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems. *Control Eng. Pract.*, 12(9):1073–1090.
- [Åström and Wittenmark, 2008] Åström, K. J. and Wittenmark, B. (2008). *Adaptive Control*. Dover Publishing, Mineola, NY, 2nd edition.
- [Babuška and Verbruggen, 1996] Babuška, R. and Verbruggen, H. (1996). An overview of fuzzy modeling for control. *Control Eng. Pract.*, 4(11):1593–1606.

- [Baert et al., 2012] Baert, A., Villez, K., and Steppe, K. (2012). Functional unfold principal component analysis for automatic plant-based stress detection in grapevine. *Funct. Plant Biol.*, 39(6):519.
- [Bingham, 1916] Bingham, E. (1916). An Investigation of the Laws of Plastic Flow. *US Bur. Stand. Bull.*, 13:309–353.
- [Birol et al., 2002] Birol, G., Ündey, C., and Çinar, A. (2002). A modular simulation package for fed-batch fermentation: penicillin production. *Comput. Chem. Eng.*, 26(11):1553–1565.
- [Bodizs et al., 2007] Bodizs, L., Titica, M., Faria, N., Srinivasan, B., Dochain, D., and Bonvin, D. (2007). Oxygen control for an industrial pilot-scale fed-batch filamentous fungal fermentation. *J. Process Control*, 17(7):595–606.
- [Bošković and Narendra, 1995] Bošković, J. D. and Narendra, K. S. (1995). Comparison of linear, nonlinear and neural-network-based adaptive controllers for a class of fed-batch fermentation processes. *Automatica*, 31(6):817–840.
- [Camacho and Picó, 2006a] Camacho, J. and Picó, J. (2006a). Multi-phase principal component analysis for batch processes modelling. *Chemom. Intell. Lab. Syst.*, 81(2):127–136.
- [Camacho and Picó, 2006b] Camacho, J. and Picó, J. (2006b). Online monitoring of batch processes using multi-phase principal component analysis. *J. Process Control*, 16(10):1021–1035.
- [Cervera et al., 2009] Cervera, A. E., Petersen, N., Lantz, A. E., Larsen, A., and Gernaey, K. V. (2009). Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. *Biotechnol. Prog.*, 25(6):1561–81.
- [Chang et al., 2016] Chang, L., Liu, X., and Henson, M. A. (2016). Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. *J. Process Control*, 42:137–149.
- [Chen and Liu, 2001] Chen, J. and Liu, J. (2001). Derivation of function space analysis based PCA control charts for batch process monitoring. *Chem. Eng. Sci.*, 56(10):3289–3304.
- [Chen et al., 2004a] Chen, L., Nguang, S. K., Chen, X. D., and Li, X. M. (2004a). Modelling and optimization of fed-batch fermentation processes using dynamic neural networks and genetic algorithms. *Biochem. Eng. J.*, 22(1):51–61.
- [Chen et al., 2004b] Chen, L. Z., Nguang, S. K., Li, X. M., and Chen, X. D. (2004b). Soft sensors for on-line biomass measurements. *Bioprocess Biosyst. Eng.*, 26(3):191–5.
- [Cherry and Fidantsef, 2003] Cherry, J. R. and Fidantsef, A. L. (2003). Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.*, 14(4):438–443.
- [Cinar et al., 2003] Cinar, A., Parulekar, S. J., Undey, C., and Birol, G. (2003). *Batch Fermentation: Modeling, Monitoring, and Control*. CRC Press.
- [Craven et al., 2014] Craven, S., Whelan, J., and Glennon, B. (2014). Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. *J. Process Control*, 24(4):344–357.
- [Cunha et al., 2002] Cunha, C. C. F., Glassey, J., Montague, G. A., Albert, S., and Mohan, P. (2002). An assessment of seed quality and its influence on productivity estimation in an industrial antibiotic fermentation. *Biotechnol. Bioeng.*, 78:658–669.

- [Dayal and MacGregor, 1997] Dayal, B. S. and MacGregor, J. F. (1997). Improved PLS algorithms. *J. Chemom.*, 11(1):73–85.
- [de Assis and Filho, 2000] de Assis, A. J. and Filho, R. M. (2000). Soft sensors development for on-line bioreactor state estimation. *Comput. Chem. Eng.*, 24:1099–1103.
- [De Leon et al., 2001] De Leon, A., Barba-de la, R., Mayani, H., Galindo, E., and Ramirez, O. T. (2001). Two useful dimensionless parameters that combine physiological, operational and bioreactor design parameters for improved control of dissolved oxygen. *Biotechnol. Lett.*, 23(13):1051–1056.
- [Deloitte, 2014] Deloitte (2014). Opportunities for the fermentation-based chemical industry. Technical report, Deloitte.
- [Dewasme et al., 2011] Dewasme, L., Srinivasan, B., Perrier, M., and Vande Wouwer, A. (2011). Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism. *J. Process Control*, 21(7):1092–1104.
- [Doan and Srinivasan, 2008] Doan, X.-T. and Srinivasan, R. (2008). Online monitoring of multiphase batch processes using phase-based multivariate statistical process control. *Comput. Chem. Eng.*, 32(1-2):230–243.
- [Doran, 1995] Doran, P. M. (1995). *Bioprocess Engineering Principles*. Academic Press, London.
- [Duan et al., 2006] Duan, S., Shi, Z., Feng, H., Duan, Z., and Mao, Z. (2006). An on-line adaptive control based on DO/pH measurements and ANN pattern recognition model for fed-batch cultivation. *Biochem. Eng. J.*, 30(1):88–96.
- [DuPont Tate & Lyle Bio Products, 2006] DuPont Tate & Lyle Bio Products (2006). \$100 million plant is world’s first to produce propanediol from corn glucose.
- [Duran-Villalobos et al., 2016] Duran-Villalobos, C. A., Lennox, B., and Lauri, D. (2016). Multivariate batch to batch optimisation of fermentation processes incorporating validity constraints. *J. Process Control*, 46:34–42.
- [Efron, 1979] Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Stat.*, 7(1):1–26.
- [Fernandes et al., 2013] Fernandes, R. L., Bodla, V. K., Carlquist, M., Heins, A. L., Lantz, A. E., Sin, G., and Gernaey, K. V. (2013). Applying mechanistic models in bioprocess development. *Adv. Biochem. Eng. Biotechnol.*, 132:137–166.
- [Ferreira et al., 2007] Ferreira, A. P., Lopes, J. a., and Menezes, J. C. (2007). Study of the application of multiway multivariate techniques to model data from an industrial fermentation process. *Anal. Chim. Acta*, 595(1-2):120–7.
- [Ferreira et al., 2001] Ferreira, L., De Souza Jr, M., and Folly, R. (2001). Development of an alcohol fermentation control system based on biosensor measurements interpreted by neural networks. *Sensors Actuators B Chem.*, 75(3):166–171.
- [Formenti et al., 2014] Formenti, L. R., Nørregaard, A., Bolic, A., Hernandez, D. Q., Hagemann, T., Heins, A.-L., Larsson, H., Mears, L., Mauricio-Iglesias, M., Krühne, U., and Gernaey, K. V. (2014). Challenges in industrial fermentation technology research. *Biotechnol. J.*, 9(6):727–38.
- [Gernaey, 2015] Gernaey, K. V. (2015). A Perspective on PSE in Fermentation Process Development and Operation. *Comput. Aided Chem. Eng.*, 37:123–130.

- [Gernaey et al., 2010] Gernaey, K. V., Lantz, A. E., Tufvesson, P., Woodley, J. M., and Sin, G. (2010). Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends Biotechnol.*, 28(7):346–54.
- [Glassey, 2013] Glassey, J. (2013). Multivariate Data Analysis for Advancing the Interpretation of Bioprocess Measurement and Monitoring Data. In *Adv Biochem Eng Biotechnol*, volume 132, pages 167–191. Springer-Verlag, Berlin Heidelberg.
- [Glassey et al., 1994] Glassey, J., Montague, G., Ward, A., and Kara, B. (1994). Enhanced supervision of recombinant *E. coli* fermentation via artificial neural networks. *Process Biochem.*, 29(5):387–398.
- [Gnoth et al., 2008] Gnoth, S., Jenzsch, M., Simutis, R., and Lübbert, A. (2008). Control of cultivation processes for recombinant protein production: a review. *Bioprocess Biosyst. Eng.*, 31(1):21–39.
- [Golabgir et al., 2015] Golabgir, A., Hoch, T., Zhariy, M., and Herwig, C. (2015). Observability analysis of biochemical process models as a valuable tool for the development of mechanistic soft sensors. *Biotechnol. Prog.*, 31(6):1703–15.
- [Gomes and Menawat, 2000] Gomes, J. and Menawat, A. S. (2000). Precise control of dissolved oxygen in bioreactors - a model-based geometric algorithm. *Chem. Eng. Sci.*, 55(1):67–78.
- [Górka-Nieć et al., 2010] Górka-Nieć, W., Perlińska-Lenart, U., Zembek, P., Palamarczyk, G., and Kruszewska, J. S. (2010). Influence of sorbitol on protein production and glycosylation and cell wall formation in *Trichoderma reesei*. *Fungal Biol.*, 114(10):855–62.
- [Grunnet Wang et al., 2013] Grunnet Wang, P., Scharling, M., Pagh Nielsen, K., Kern-Hansen, C., and Wittchen, K. B. (2013). 2001-2010 Danish Design Reference Year. Technical report, Danish Meteorological Institute.
- [Gurden et al., 2001] Gurden, S. P., Westerhuis, J. A., Bro, R., and Smilde, A. K. (2001). A comparison of multiway regression and scaling methods. *Chemom. Intell. Lab. Syst.*, 59(1-2):121–136.
- [Hastie et al., 2009] Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning - Data Mining, Inference, and Prediction, Second Edition*.
- [Helton and Davis, 2003] Helton, J. and Davis, F. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab. Eng. Syst. Saf.*, 81(1):23–69.
- [Henes and Sonnleitner, 2007] Henes, B. and Sonnleitner, B. (2007). Controlled fed-batch by tracking the maximal culture capacity. *J. Biotechnol.*, 132(2):118–26.
- [Hisbullah and Ramachandran, 2002] Hisbullah, M. H. and Ramachandran, K. (2002). Comparative evaluation of various control schemes for fed-batch fermentation. *Bioprocess Biosyst. Eng.*, 24(5):309–318.
- [Hisbullah and Ramachandran, 2003] Hisbullah, M. H. and Ramachandran, K. (2003). Design of a Fuzzy Logic Controller for Regulating Substrate Feed to Fed-Batch Fermentation. *Food Bioprod. Process.*, 81(2):138–146.
- [Hofmeester, 1988] Hofmeester, J. J. (1988). Gas hold-up measurements in bioreactors. *Trends Biotechnol.*, 6(1):19–22.

- [Holland, 1984] Holland, J. H. (1984). Genetic Algorithms and Adaptation. In *Adapt. Control Ill-Defined Syst.*, pages 317–333. Springer US, Boston, MA.
- [Horiuchi and Hiraga, 1999] Horiuchi, J.-I. and Hiraga, K. (1999). Industrial application of fuzzy control to large-scale recombinant vitamin B2 production. *J. Biosci. Bioeng.*, 87(3):365–371.
- [Hui et al., 2002] Hui, J. P. M., White, T. C., and Thibault, P. (2002). Identification of glycan structure and glycosylation sites in cellobiohydrolase II and endoglucanases I and II from *Trichoderma reesei*. *Glycobiology*, 12(12):837–849.
- [Iman and Conover, 1982] Iman, R. L. and Conover, W. J. (1982). A distribution-free approach to inducing rank correlation among input variables. *Commun. Stat. - Simul. Comput.*, 11(3):311–334.
- [Jenzsch et al., 2006a] Jenzsch, M., Gnoth, S., Beck, M., Kleinschmidt, M., Simutis, R., and Lübbert, A. (2006a). Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes. *J. Biotechnol.*, 127(1):84–94.
- [Jenzsch et al., 2006b] Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., and Lübbert, A. (2006b). Improving the batch-to-batch reproducibility in microbial cultures during recombinant protein production by guiding the process along a predefined total biomass profile. *Bioprocess Biosyst. Eng.*, 29(5-6):315–21.
- [Jenzsch et al., 2007] Jenzsch, M., Gnoth, S., Kleinschmidt, M., Simutis, R., and Lübbert, A. (2007). Improving the batch-to-batch reproducibility of microbial cultures during recombinant protein production by regulation of the total carbon dioxide production. *J. Biotechnol.*, 128(4):858–67.
- [Jenzsch et al., 2006c] Jenzsch, M., Simutis, R., and Luebbert, A. (2006c). Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. *J. Biotechnol.*, 122(4):483–93.
- [Johnsson et al., 2013] Johnsson, O., Andersson, J., Lidén, G., Johnsson, C., and Hägglund, T. (2013). Feed rate control in fed-batch fermentations based on frequency content analysis. *Biotechnol. Prog.*, 29(3):817–24.
- [Kassidas et al., 1998] Kassidas, A., MacGregor, J. F., and Taylor, P. A. (1998). Synchronization of batch trajectories using dynamic time warping. *AIChE J.*, 44(4):864–875.
- [Keogh and Ratanamahatana, 2004] Keogh, E. and Ratanamahatana, C. A. (2004). Exact indexing of dynamic time warping. *Knowl. Inf. Syst.*, 7(3):358–386.
- [Kovárová-Kovar et al., 2000] Kovárová-Kovar, K., Gehlen, S., Kunze, A., Keller, T., von Däniken, R., Kolb, M., and van Loon, A. P. (2000). Application of model-predictive control based on artificial neural networks to optimize the fed-batch process for riboflavin production. *J. Biotechnol.*, 79(1):39–52.
- [Kuprijanov et al., 2013] Kuprijanov, A., Schaepe, S., Simutis, R., and Lübbert, A. (2013). Model predictive control made accessible to professional automation systems in fermentation technology. *Biosyst. Inf. Technol.*, 2(2):26–31.
- [Landau et al., 2011] Landau, I. D., Lozano, R., M’Saad, M., and Karimi, A. (2011). *Adaptive Control*. Communications and Control Engineering. Springer-Verlag, London.
- [Latour et al., 1986] Latour, P., Sharpe, J., and Delaney, M. (1986). Estimating Benefits from Advanced Control. *ISA Trans.*, 25(4):13–21.

- [Laurí et al., 2014] Laurí, D., Lennox, B., and Camacho, J. (2014). Model predictive control for batch processes: Ensuring validity of predictions. *J. Process Control*, 24(1):239–249.
- [Le et al., 2012] Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., and Hu, W.-S. (2012). Multivariate analysis of cell culture bioprocess data—lactate consumption as process indicator. *J. Biotechnol.*, 162(2-3):210–23.
- [Lee et al., 1999] Lee, J., Lee, S. Y., Park, S., and Middelberg, A. P. J. (1999). Control of fed-batch fermentations. *Biotechnol. Adv.*, 17(1):29–48.
- [Lee and Kim, 2015] Lee, S. Y. and Kim, H. U. (2015). Systems strategies for developing industrial microbial strains. *Nat. Biotechnol.*, 33(10):1061–1072.
- [Lennox et al., 2000] Lennox, B., Hiden, H., Montague, G., Kornfeld, G., and Goulding, P. (2000). Application of multivariate statistical process control to batch operations. *Comput. Chem. Eng.*, 24(2-7):291–296.
- [Linko et al., 1999] Linko, S., Zhu, Y.-H., and Linko, P. (1999). Applying neural networks as software sensors for enzyme engineering. *Trends Biotechnol.*, 17(4):155–162.
- [Long et al., 2014] Long, Q., Liu, X., Yang, Y., Li, L., Harvey, L., McNeil, B., and Bai, Z. (2014). The development and application of high throughput cultivation technology in bioprocess development. *J. Biotechnol.*, 192:323–338.
- [Louwerse and Smilde, 2000] Louwerse, D. and Smilde, A. (2000). Multivariate statistical process control of batch processes based on three-way models. *Chem. Eng. Sci.*, 55(7):1225–1235.
- [Lübbert and Jørgensen, 2001] Lübbert, A. and Jørgensen, S. B. (2001). Bioreactor performance: a more scientific approach for practice. *J. Biotechnol.*, 85(2):187–212.
- [Lübbert and Simutis, 1994] Lübbert, A. and Simutis, R. (1994). Using measurement data in bioprocess modelling and control. *Trends Biotechnol.*, 12(8):304–311.
- [Luttmann et al., 2012] Luttmann, R., Bracewell, D. G., Cornelissen, G., Gernaey, K. V., Glassey, J., Hass, V. C., Kaiser, C., Preusse, C., Striedner, G., and Mandenius, C.-F. (2012). Soft sensors in bioprocessing: a status report and recommendations. *Biotechnol. J.*, 7(8):1040–8.
- [Mears et al., 2016] Mears, L., Nørregård, R., Sin, G., Gernaey, K., Stocks, S., Albaek, M., and Villez, K. (2016). Functional unfold principal component regression methodology for analysis of industrial batch process data. *AIChE J.*, 62(6):1986–1994.
- [Mercier et al., 2013] Mercier, S. M., Diepenbroek, B., Dalm, M. C. F., Wijffels, R. H., and Streefland, M. (2013). Multivariate data analysis as a PAT tool for early bioprocess development data. *J. Biotechnol.*, 167(3):262–70.
- [Metropolis and Ulam, 1949] Metropolis, N. and Ulam, S. (1949). The Monte Carlo Method. *J. Am. Stat. Assoc.*, 44(247):335–341.
- [Metzner and Otto, 1957] Metzner, a. B. and Otto, R. E. (1957). Agitation of non-Newtonian fluids. *AIChE J.*, 3(1):3–10.
- [Montague et al., 1989] Montague, G., Morris, A., and Ward, A. (1989). Fermentation monitoring and control: a perspective. *Biotechnol. Genet. Eng. Rev.*, 7(1):147–188.
- [Neubauer and Junne, 2010] Neubauer, P. and Junne, S. (2010). Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr. Opin. Biotechnol.*, 21(1):114–21.

- [Nomikos and MacGregor, 1994] Nomikos, P. and MacGregor, J. F. (1994). Monitoring batch processes using multiway principal component analysis. *AIChE J.*, 40(8):1361–1375.
- [Nomikos and MacGregor, 1995] Nomikos, P. and MacGregor, J. F. (1995). Multi-way partial least squares in monitoring batch processes. *Chemom. Intell. Lab. Syst.*, 30(1):97–108.
- [Novozymes A/S, a] Novozymes A/S. Novozymes honored by the UN.
- [Novozymes A/S, b] Novozymes A/S. World leader in biological solutions.
- [Oliveira et al., 2005] Oliveira, R., Clemente, J. J., Cunha, A. E., and Carrondo, M. J. T. (2005). Adaptive dissolved oxygen control through the glycerol feeding in a recombinant *Pichia pastoris* cultivation in conditions of oxygen transfer limitation. *J. Biotechnol.*, 116(1):35–50.
- [Oliveira et al., 2004] Oliveira, R., Simutis, R., and Feyo de Azevedo, S. (2004). Design of a stable adaptive controller for driving aerobic fermentation processes near maximum oxygen transfer capacity. *J. Process Control*, 14(6):617–626.
- [Olsvik and Kristiansen, 1994] Olsvik, E. and Kristiansen, B. (1994). Rheology of filamentous fermentations. *Biotechnol. Adv.*, 12(1):1–39.
- [Peng et al., 2013] Peng, J., Meng, F., and Ai, Y. (2013). Time-dependent fermentation control strategies for enhancing synthesis of marine bacteriocin 1701 using artificial neural network and genetic algorithm. *Bioresour. Technol.*, 138:345–52.
- [Posch et al., 2013] Posch, A. E., Herwig, C., and Spadiut, O. (2013). Science-based bioprocess design for filamentous fungi. *Trends Biotechnol.*, 31(1):37–44.
- [Qin and Badgwell, 2003] Qin, S. J. and Badgwell, T. A. (2003). A survey of industrial model predictive control technology. *Control Eng. Pract.*, 11:733–764.
- [Quintanilla et al., 2015] Quintanilla, D., Hagemann, T., Hansen, K., and Gernaey, K. V. (2015). Fungal Morphology in Industrial Enzyme Production - Modelling and Monitoring. In *Filaments in Bioprocesses*, volume 123, pages 29–54. Springer International Publishing.
- [Ramsey and Silverman, 2005] Ramsey, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag.
- [Riisgaard and Andersson, 2013] Riisgaard, F. K. and Andersson, J. (2013). Advanced fermentation control.
- [Riley et al., 2000] Riley, G. L., Tucker, K. G., Paul, G. C., and Thomas, C. R. (2000). Effect of biomass concentration and mycelial morphology on fermentation broth rheology. *Biotechnol. Bioeng.*, 68(2):160–72.
- [Roels and Heijnen, 1980] Roels, J. A. and Heijnen, J. J. (1980). Power dissipation and heat production in bubble columns: Approach based on nonequilibrium thermodynamics. *Biotechnol. Bioeng.*, 22(11):2399–2404.
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- [Sagmeister et al., 2013] Sagmeister, P., Wechselberger, P., Jazini, M., Meitz, A., Langemann, T., and Herwig, C. (2013). Soft sensor assisted dynamic bioprocess control: Efficient tools for bioprocess development. *Chem. Eng. Sci.*, 96:190–198.

- [Santos et al., 2012] Santos, L., Dewasme, L., Coutinho, D., and Wouwer, A. V. (2012). Non-linear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. *Comput. Chem. Eng.*, 39:143–151.
- [Schuster and Schmoll, 2010] Schuster, A. and Schmoll, M. (2010). Biology and biotechnology of *Trichoderma*. *Appl. Microbiol. Biotechnol.*, 87(3):787–99.
- [Seborg et al., 2003] Seborg, D. E., Edgar, T. F., and Mellichamp, D. A. (2003). *Process Dynamics and Control*. 2nd edition.
- [Sin et al., 2009] Sin, G., Gernaey, K. V., and Lantz, A. E. (2009). Good Modeling Practice for PAT Applications : Propagation of Input Uncertainty and Sensitivity Analysis. *Biotechnol. Prog.*, 25(4):1043–1053.
- [Smets et al., 2004] Smets, I. Y., Claes, J. E., November, E. J., Bastin, G. P., and Van Impe, J. F. (2004). Optimal adaptive control of (bio)chemical reactors: past, present and future. *J. Process Control*, 14(7):795–805.
- [Sonnleitner, 2013] Sonnleitner, B. (2013). Automated Measurement and Monitoring of Bioprocesses : Key Elements of the M3C Strategy. *Adv. Biochem. Eng. Biotechnol.*, 132:1–33.
- [Soons et al., 2006] Soons, Z. I. T. A., Voogt, J. A., van Straten, G., and van Boxtel, A. J. B. (2006). Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control. *J. Biotechnol.*, 125(2):252–68.
- [Stanke and Hitzmann, 2013] Stanke, M. and Hitzmann, B. (2013). Automatic control of bioprocesses. In *Meas. Monit. Model. Control Bioprocesses*, volume 132, pages 35–63. Springer-Verlag, Berlin, Heidelberg.
- [Stocks, 2013] Stocks, S. (2013). Industrial enzyme production for the food and beverage industries: process scale up and scale down. In McNeil, B., Archer, D., Giavasis, I., and Harvey, L., editors, *Microb. Prod. Food Ingredients, Enzym. Nutraceuticals*, pages 144–172. Woodhead Publishing Limited, Cambridge.
- [Sundström and Enfors, 2008] Sundström, H. and Enfors, S. O. (2008). Software sensors for fermentation processes. *Bioprocess Biosyst. Eng.*, 31(2):145–152.
- [The MathWorks Inc., 2013] The MathWorks Inc. (2013). Matlab 8.2 (R2013b).
- [Thornhill et al., 2004] Thornhill, N. F., Shoukat Choudhury, M., and Shah, S. L. (2004). The impact of compression on data-driven process analyses. *J. Process Control*, 14(4):389–398.
- [Tolan and Foody, 1999] Tolan, J. S. and Foody, B. (1999). Cellulase from Submerged Fermentation. In Tsao, G. T., editor, *Recent Prog. Bioconversion Lignocellul.*, volume 65 of *Advances in Biochemical Engineering/Biotechnology*, pages 41–67. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Ündey and Cinar, 2002] Ündey, C. and Cinar, A. (2002). Statistical monitoring of multistage, multiphase batch processes. *IEEE Control Syst. Mag.*, 22(5):40–52.
- [U.S. Food and Drug Administration, 2015] U.S. Food and Drug Administration (2015). Microorganisms and Microbial-Derived Ingredients Used in Food (Partial List).
- [Velut et al., 2007] Velut, S., de Maré, L., and Hagander, P. (2007). Bioreactor control using a probing feeding strategy and mid-ranging control. *Control Eng. Pract.*, 15(2):135–147.
- [Villadsen et al., 2011] Villadsen, J., Nielsen, J., and Lidén, G. (2011). *Bioreaction Engineering Principles*. Springer, New York, third edition.

- [Villez et al., 2008] Villez, K., Vanrolleghem, P. A., Rosén, C., Anctil, F., and Duchesne, C. (2008). Qualitative representation of trends: an alternative approach to process diagnosis and control. *Water Sci. Technol.*, 57(10):1525–32.
- [von Stosch et al., 2014] von Stosch, M., Oliveira, R., Peres, J., and Foyo de Azevedo, S. (2014). Hybrid semi-parametric modeling in process systems engineering: Past, present and future. *Comput. Chem. Eng.*, 60:86–101.
- [Waites et al., 2001] Waites, M. J., Morgan, N. L., Rockey, J. S., and Higton, G. (2001). *Industrial microbiology*. Blackwell Science.
- [Wan et al., 2014] Wan, J., Marjanovic, O., and Lennox, B. (2014). Uneven batch data alignment with application to the control of batch end-product quality. *ISA Trans.*, 53(2):584–90.
- [Wechselberger et al., 2012] Wechselberger, P., Sagmeister, P., Engelking, H., Schmidt, T., Wenger, J., and Herwig, C. (2012). Efficient feeding profile optimization for recombinant protein production using physiological information. *Bioprocess Biosyst. Eng.*, 35(9):1637–49.
- [Yuan et al., 2014] Yuan, X., Ge, Z., and Song, Z. (2014). Locally Weighted Kernel Principal Component Regression Model for Soft Sensing of Nonlinear Time-Variant Processes. *Ind. Eng. Chem. Res.*, 53(35):13736–13749.
- [Zeng et al., 1993] Zeng, F., Dahhou, B., and Nihtila, M. (1993). Adaptive control of a nonlinear fermentation process via MRAC technique. *Appl. Math. Model.*, 17(2):58–69.
- [Zhang and Lennox, 2004] Zhang, H. and Lennox, B. (2004). Integrated condition monitoring and control of fed-batch fermentation processes. *J. Process Control*, 14(1):41–50.
- [Zhang et al., 1994] Zhang, X.-C., Visala, A., Halme, A., and Linko, P. (1994). Functional state modeling and fuzzy control of fed-batch aerobic baker’s yeast process. *J. Biotechnol.*, 37(1):1–10.

CAPEC-PROCESS Research Center
Department of Chemical and Biochemical Engineering
Technical University of Denmark
Building 229
DK - 2800 Kgs. Lyngby
Denmark

Phone: +45 45 25 28 00
Web: capec-process.kt.dtu.dk